

ĐẠI HỌC QUỐC GIA HÀ NỘI  
NGUYỄN QUANG BẦU

# LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC



NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

NGUYỄN QUANG BÁU

LÝ THUYẾT XÁC SUẤT  
VÀ THỐNG KÊ TOÁN HỌC  
(In lần thứ ba)

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI



## LỜI NÓI ĐẦU

Giáo trình “*Lý thuyết xác suất và thống kê toán*” được biên soạn trên cơ sở các bài giảng của tác giả trình bày nhiều năm (có tham khảo các tài liệu trong và ngoài nước [1-6]) và theo chương trình toán giai đoạn 2 dành cho sinh viên ngành học Vật lý, Khoa học Vật liệu, Khoa học và Công nghệ hạt nhân, Vô tuyến-Điện tử của Trường Đại học Tổng hợp Hà Nội (nay là Trường Đại học Khoa học Tự nhiên thuộc Đại học Quốc gia Hà Nội). Giáo trình cũng rất bổ ích và có thể dùng làm tài liệu tham khảo, giảng dạy cho sinh viên các trường Đại học Sư phạm, Đại học Kỹ thuật, Đại học Kinh tế Quốc dân và các trường khác trong nước có học môn “Lý thuyết xác suất và thống kê toán”. Khi viết chúng tôi đã cố gắng đạt tới mục đích đề ra là cơ bản, hiện đại và Việt Nam.

Giáo trình trình bày những khái niệm cơ bản của lý thuyết xác suất (chương 1); một số công thức tính xác suất quan trọng trong lý thuyết xác suất (chương 2); những khái niệm về đại lượng ngẫu nhiên và các tính chất, tham số đặc trưng của chúng (chương 3); những khái niệm về thống kê toán như mẫu thống kê và ước lượng tham số (chương 4); kiểm nghiệm các giả thiết thống kê (chương 5). Ngoài phần bài giảng, giáo trình có đưa ra phần phụ lục (bổ túc toán và các bảng tra cứu thường dùng); một lượng lớn bài tập có kèm theo lời giải và chỉ dẫn rõ ràng, cụ thể. Những bài tập này nhằm giúp cho học viên dễ nắm bắt và hiểu sâu sắc hơn nội dung bài giảng và đặc biệt có tác dụng rèn luyện kỹ năng vận dụng lý thuyết xác suất và thống kê toán trong các ngành Khoa học Vật lý, Khoa học Vật liệu, Khoa học và Công nghệ hạt nhân, Vô tuyến-Điện tử, ... cũng như trong các vấn đề thực tế của kinh tế - xã hội đặt ra.

Giáo trình là một trong các giáo trình thuộc “Bộ giáo trình toán cho Vật lý” do các GS, PGS, TSKH, TS, cán bộ giảng dạy lâu năm của Bộ môn Vật lý lý thuyết, Khoa Vật lý, Trường Đại học Tổng hợp Hà Nội (nay là Trường Đại học Khoa học Tự nhiên thuộc Đại học Quốc gia Hà Nội) đảm nhận giảng dạy và biên soạn.

Viết giáo trình “Lý thuyết xác suất và Thống kê toán” là một việc rất khó khăn, vì vậy mặc dù đã hết sức cố gắng, nhưng giáo trình chắc không tránh khỏi một số thiếu sót. Chúng tôi hy vọng nhận được nhiều ý kiến nhận xét, đóng góp của bạn đọc để giáo trình ngày càng hoàn thiện hơn.

Chúng tôi xin cảm ơn các bạn đồng nghiệp thuộc bộ môn Vật lý lý thuyết, khoa Vật lý, Trường Đại học Khoa học Tự nhiên thuộc Đại học Quốc gia Hà Nội, các GS.TS Nguyễn Văn Thỏ, GS.TSKH Nguyễn Xuân Hãn, PGS.TSKH Nguyễn Văn Hùng, PGS.TS Lê Văn Trục, PGS Phạm Công Dũng, PGS.TS Phạm Tế Thế đã đóng góp nhiều ý kiến quý báu trong quá trình biên soạn giáo trình này.

**Tác giả**

**GS.TS Nguyễn Quang Báu**



## *Chương 1*

# **NHỮNG KHÁI NIỆM CƠ BẢN CỦA LÝ THUYẾT XÁC SUẤT**

### **1.1. KHÁI NIỆM VỀ PHÉP THỬ, BIẾN CỐ VÀ XÁC SUẤT CỦA BIẾN CỐ**

#### **1.1.1. Phép thử và biến cố**

Các khái niệm đầu tiên được gặp trong lý thuyết xác suất là khái niệm về phép thử và biến cố. “Phép thử” được hiểu là thực hiện một nhóm các điều kiện xác định nào đó, nó có thể là một thí nghiệm hay một việc quan sát sự xuất hiện một hiện tượng nào đó. Một phép thử có thể có nhiều kết cục khác nhau, các kết cục này được gọi là các “biến cố”. Ta qui ước ký hiệu các biến cố bằng các chữ in A, B, C...(và đôi khi có thể kèm theo chỉ số). Ví dụ:

- Gieo một con xúc xắc là phép thử. Phép thử này có thể kết thúc bằng một trong 6 kết cục sau: xuất hiện mặt 1 chấm, 2 chấm, 3 chấm, 4 chấm, 5 chấm, 6 chấm. Mỗi kết cục này cũng như các kết cục phức hợp của chúng: xuất hiện số chấm chẵn, số chấm lẻ,... đều là những biến cố.

- Quan sát trạng thái hoạt động của một máy móc là phép thử. Hai kết cục: máy chạy tốt và máy hỏng hóc là hai biến cố.

Tùy theo tính chất xuất hiện của các biến cố trong phép thử mà ta chia chúng làm 3 loại: biến cố chắc chắn, biến cố không thể có và biến cố ngẫu nhiên.

Biến cố chắc chắn là biến cố nhất thiết xảy ra khi thực hiện phép thử.

Biến cố không thể có là biến cố nhất thiết không xảy ra khi thực hiện phép thử.

Biến cố ngẫu nhiên là biến cố có thể xảy ra khi thực hiện phép thử.

Ta quy ước ký hiệu biến cố chắc chắn là chữ in U, biến cố không thể có là chữ in V. Ví dụ:

- Xét phép thử: gieo một con xúc xắc. Thế thì biến cố xuất hiện số chấm nhỏ hơn hoặc bằng 6 là biến cố chắc chắn, biến cố xuất hiện số chấm lớn hơn 6 là biến cố không thể có, biến cố xuất hiện số chấm bằng 2 là biến cố ngẫu nhiên.

- Xét phép thử: bắn một phát đạn vào bia. Thế thì biến cố trúng hoặc trượt bia là biến cố chắc chắn, biến cố trúng bia 2 lần là biến cố không thể có, biến cố trúng bia là biến cố ngẫu nhiên.

### **1.1.2. Xác suất của biến cố**

Quan sát các biến cố ngẫu nhiên, ta thấy rằng khả năng xuất hiện của chúng nói chung không đồng đều, một số biến cố thường hay xảy ra, một số khác lại thường ít xảy ra. Từ đó nảy sinh vấn đề tìm cách “đo lường” khả năng xuất hiện của một biến cố. Để làm điều đó, người ta đã gán cho mỗi biến cố một số không âm và số này được gọi là xác suất của biến cố.

Ta qui ước ký hiệu xác suất của biến cố A là  $P(A)$ . Xác suất  $P(A)$  của một biến cố A phải xây dựng sao cho thoả mãn các đòi hỏi hợp lý sau:

- Xác suất của biến cố chắc chắn U bằng 1:  $P(U)=1$  (vì với biến cố chắc chắn thì 100% xảy ra).

- Xác suất của biến cố không thể có V bằng 0:  $P(V)=0$  (vì với biến cố không thể có thì 100% không xảy ra ).

- Xác suất của biến cố ngẫu nhiên A bị kẹp giữa số 0 và số 1:  $0 < P(A) < 1$ .



## 1.2. CÁC ĐỊNH NGHĨA VỀ XÁC SUẤT CỦA BIẾN CỐ

### 1.2.1. Định nghĩa về xác suất theo quan điểm cổ điển

#### a) Định nghĩa

Giả sử một phép thử có tất cả  $n$  kết cục đồng khả năng, trong đó có  $m$  kết cục thuận lợi cho biến cố  $A$  xuất hiện. Khi đó xác suất của biến cố  $A$  là tỷ số kết cục thuận lợi cho  $A$  xuất hiện trên số kết cục đồng khả năng:

$$P(A) = \frac{(\text{số kết cục thuận lợi cho } A \text{ xuất hiện})}{(\text{số kết cục đồng khả năng})} = \frac{m}{n} \quad (1)$$

Từ công thức trên dễ dàng suy ra các tính chất chung của  $P(A)$ :

- Đối với biến cố chắc chắn  $U$  thì mọi kết cục đồng khả năng của phép thử đều là các kết cục thuận lợi cho biến cố đó xuất hiện  $m=n$ , do đó  $P(U) = \frac{n}{n} = 1$ .

- Đối với biến cố không thể có  $V$  thì không có kết cục nào thuận lợi cho biến cố đó xuất hiện nên  $m=0$ , do đó  $P(V) = \frac{0}{n} = 0$ .

- Đối với biến cố ngẫu nhiên  $A$  thì số kết cục thuận lợi cho  $A$  xuất hiện bị kẹp giữa số 0 và số  $n$  nên  $0 < m < n$ , do đó  $0 < P(A) = \frac{m}{n} < 1$ .

Cách tính xác suất dựa trên định nghĩa xác suất theo quan điểm cổ điển có ưu điểm là đơn giản và trực quan, nhưng có hạn chế là phạm vi sử dụng của nó không lớn, chỉ dành cho loại phép thử gồm một số hữu hạn các kết cục và mọi kết cục đều có cùng một khả năng xuất hiện mà thôi.

#### b) Các ví dụ

- Ví dụ 1: Gieo một con xúc xắc hoàn toàn đối xứng. Hãy tính xác suất để: a) Được mặt sáu chấm?; b) Được mặt có số chấm chẵn?

*Giải:* Gọi  $A$  là biến cố được mặt sáu chấm,  $B$  là biến cố được mặt có số chấm chẵn. Ở đây số kết cục đồng khả năng  $n=6$ , số kết cục

thuận lợi cho A xuất hiện  $m_A=1$ , số kết cục thuận lợi cho B xuất hiện  $m_B=3$ . Do đó ta có :

$$P(A) = \frac{1}{6} ; P(B) = \frac{3}{6} = \frac{1}{2}$$

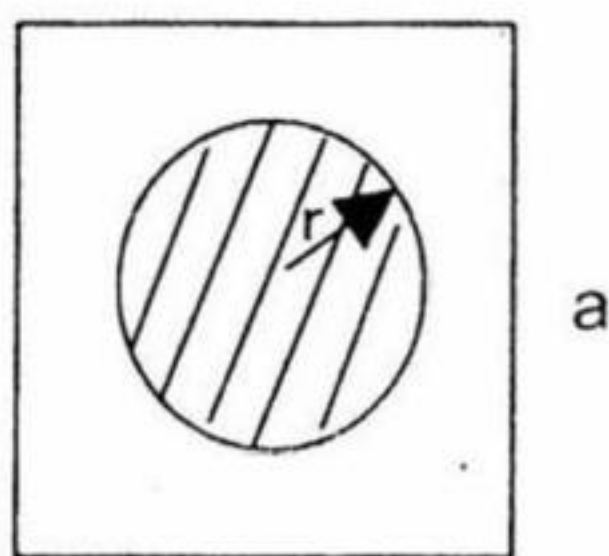
- Ví dụ 2: Gieo hai con xúc xắc hoàn toàn đối xứng. Hãy tính xác suất để được hai mặt sáu chấm .

*Giải:* Ký hiệu a, b là số chấm trên mặt con xúc xắc thứ nhất và thứ hai ( $1 \leq a \leq 6 ; 1 \leq b \leq 6$  ). Như vậy, kết quả có được số chấm khi gieo hai con xúc xắc có thể ký hiệu bằng bộ hai số (a,b). Gọi c là biến cố được hai mặt sáu chấm. Theo qui tắc kết hợp nhân số kết cục đồng khả năng bằng số bộ hai số (a,b), tức  $n=6 \times 6=36$ . Số kết cục thuận lợi cho c xuất hiện ứng với trường hợp  $a=b=6$ , tức  $m_c=1$ . Do đó ta có:

$$P(c) = \frac{1}{36}$$

### c) Hạn chế

Định nghĩa xác suất theo quan điểm cổ điển hạn chế là chỉ áp dụng cho các kết cục đồng khả năng và các số m, n trong công thức (1) phải được xác định cụ thể và hữu hạn. Trong nhiều trường hợp các điều kiện này rất khó thực hiện hoặc không thể thực hiện được. Ví dụ, để tính xác suất sinh con trai không thể cho  $m=1$  và  $n=2$  rồi theo công thức (1) để tính; hoặc khi gieo một chấm điểm vào một hình vuông cạnh a, bên trong có một hình tròn bán kính r (hình 1), ta cũng không thể tính xác suất để chấm điểm rơi vào hình tròn theo công thức (1) vì không biết m bằng bao nhiêu, và n bằng bao nhiêu. Các hạn chế này sẽ được bù lấp bởi các định nghĩa về xác suất theo quan điểm thống kê và theo quan điểm hình học.



Hình 1



### 1.2.2. Định nghĩa về xác suất theo quan điểm thống kê

#### a) Khái niệm về tần suất xuất hiện biến cố A (f)

Có thể coi tần suất xuất hiện biến cố A là tỷ số giữa số x phép thử trong đó có A xuất hiện và tổng số n phép thử được thực hiện. Nếu ký hiệu f là tần suất xuất hiện biến cố A, thì ta có:

$$f = \frac{\text{Số phép thử có A}}{\text{Tổng số phép thử được thực hiện}} = \frac{x}{n} \quad (2)$$

*Ví dụ:* Trong 1000 trẻ sơ sinh trong tháng 5 năm 1997 ở Hà Nội có 525 trai và 475 gái. Như vậy tần suất xuất hiện con gái ở đây là:

$$f = \frac{475}{1000} = 0,475$$

#### b) Định nghĩa

Tần suất xuất hiện biến cố A trong n phép thử luôn luôn dao động xung quanh một số không đổi P(A), và khi n tiến tới vô hạn thì tần suất xuất hiện biến cố A càng gần đến số không đổi P(A) đó. Khi đó số P(A) được gọi là xác suất của biến cố A theo quan điểm thống kê. Nói cách khác, ta có :

$$P(A) = \lim_{n \rightarrow \infty} f = \lim_{n \rightarrow \infty} \frac{x}{n} \quad (3)$$

Từ định nghĩa xác suất theo quan điểm thống kê nói trên, cũng dễ dàng suy ra các tính chất chung của P(A):

- Đối với biến cố chắc chắn U thì tần suất xuất hiện biến cố U:

$$f = \frac{n}{n} = 1 \text{ . Dẫn đến xác suất } P(U) = \lim_{n \rightarrow \infty} f = 1 \text{ .}$$

- Đối với biến cố không thể có V thì tần suất xuất hiện biến cố V:

$$f = \frac{0}{n} = 0 \text{ . Dẫn đến xác suất } P(V) = \lim_{n \rightarrow \infty} f = 0 \text{ .}$$

- Đối với biến cố ngẫu nhiên A thì tần suất xuất hiện biến cố A:  
 $f = \frac{x}{n}$  (với  $0 < x < n$  và  $0 < f < 1$ ). Dẫn đến xác suất  $0 < P(A) = \lim_{n \rightarrow \infty} f < 1$ .

### *c) Ứng dụng*

Trong thực tế khi ứng dụng định nghĩa xác suất theo quan điểm thống kê ta không thể thực hiện số phép thử lớn vô hạn được và không thể tính chính xác xác suất biến cố A theo công thức (3) được mà người ta thường lấy giá trị của tần suất xuất hiện biến cố A trong một loạt khá lớn các phép thử làm giá trị gần đúng của xác suất  $P(A)$ ; phương pháp xác định xác suất theo quan điểm thống kê được áp dụng có hiệu quả trong việc tìm ra qui luật diễn biến phức tạp về thời tiết, về tỷ lệ phế phẩm, truyền tin qua các tầng điện ly, lập kích thước quần áo may sẵn, nghiên cứu công hiệu của thuốc men, trong nhân chủng học, xã hội học... .

### **1.2.3. Định nghĩa xác suất theo quan điểm hình học**

Định nghĩa xác suất theo quan điểm thống kê khắc phục được hạn chế của định nghĩa xác suất theo quan điểm cổ điển về đòi hỏi các kết cục của phép thử phải đồng khả năng xuất hiện. Để khắc phục hạn chế của định nghĩa xác suất theo quan điểm cổ điển về đòi hỏi số kết cục của phép thử xác định cụ thể và hữu hạn (đồng thời vẫn giả thiết các kết cục đồng khả năng) người ta đưa ra định nghĩa xác suất theo quan điểm hình học.

#### *a) Định nghĩa*

Xét một phép thử có vô hạn các kết cục đồng khả năng. Giả sử ta có thể biểu diễn tập hợp mọi kết cục này bởi một miền hình học G nào đó (một đoạn thẳng, một miền phẳng, một mảnh mặt cong hay một khối không gian) và những kết cục thuận lợi cho biến cố A xuất hiện



bởi một miền hình học con  $g$  thuộc  $G$ . Với giả thiết trên xác suất của biến cố  $A$  được tính như là tỷ số giữa "kích thước" miền  $g$  trên "kích thước" miền  $G$ , tức là:

$$P(A) = \frac{\text{"kích thước" miền } g}{\text{"kích thước" miền } G} \quad (4)$$

Tùy theo  $g$  và  $G$  là đoạn thẳng, miền phẳng (cong) hay khối không gian mà kích thước được hiểu là độ dài, diện tích hay thể tích.

#### *b) Các ví dụ*

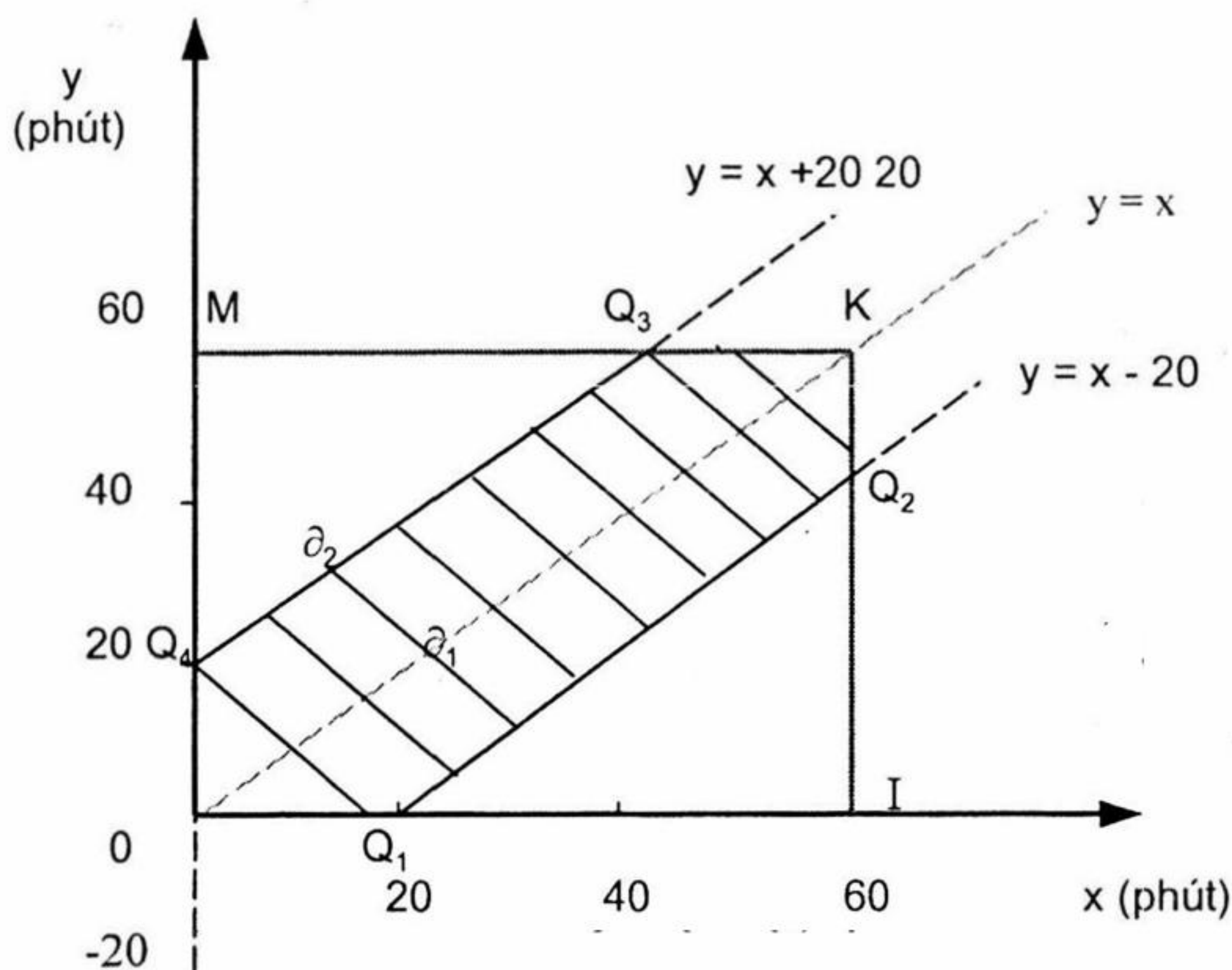
- *Ví dụ 1:* Đâm một mũi kim một cách ngẫu nhiên vào hình vuông cạnh  $a$ , trong có hình tròn nội tiếp bán kính  $a/2$ . Hãy tìm xác suất mũi kim rơi vào hình tròn.

*Giải:* Gọi  $A$  là biến cố mũi kim rơi vào hình tròn. Khi đó xác suất biến cố mũi kim rơi vào hình tròn được tính theo công thức (4) và bằng:

$$P(A) = \frac{\text{diện tích hình tròn}}{\text{diện tích hình vuông}} = \frac{\pi \frac{a^2}{4}}{a^2} = \frac{\pi}{4}$$

- *Ví dụ 2:* Hai người A và B hẹn gặp nhau tại một địa điểm xác định trong vòng từ 0 giờ đến 1 giờ. Người đến trước chờ người kia quá 20 phút thì sẽ bỏ đi. Hãy tính xác suất họ gặp được nhau, biết rằng mỗi người có thể đến chỗ hẹn vào một thời điểm bất kỳ trong khoảng thời gian trên.

*Giải:* Gọi  $x$  là lúc người A đến điểm hẹn,  $y$  là lúc người B đến điểm hẹn ( $x, y$  tính ra phút). Mọi kết cục đồng khả năng là mọi cặp số  $(x, y)$  với điều kiện  $0 \leq x \leq 60, 0 \leq y \leq 60$ . Tập hợp này có thể biểu diễn hình học bởi hình vuông OIKM (hình 2).



**Hình 2**

Các kết cục thuận lợi cho hai người gặp nhau là cặp số  $(x,y)$  sao cho:

$$|x - y| \leq 20$$

Trên hình vẽ tập hợp các cặp số này ứng với miền con của hình vuông gồm giữa các đường thẳng  $y=x+20$  và  $y=x-20$  (phần gạch gạch trong hình 2). Gọi  $C$  là biến cố hai người gặp được nhau. Khi đó  $P(C)$  tính theo công thức (4) bằng:

$$P(C) = \frac{\text{diện tích } OQ_1Q_2KQ_3Q_4O}{\text{diện tích } (OIKMO)} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$$

### 1.3. MỐI QUAN HỆ VÀ CÁC PHÉP TÍNH GIỮA CÁC BIẾN CỐ

Trong thực tế ta thường gặp các loại biến cố phức tạp do vậy việc tính trực tiếp xác suất của chúng đôi khi gặp khó khăn. Lúc đó, người ta có thể tính gián tiếp xác suất của chúng thông qua các mối quan hệ giữa các biến cố và các phép tính giữa chúng.



### 1.3.1. Mối quan hệ giữa các biến cố

a) Ta gọi biến cố A kéo theo biến cố B, nếu A xuất hiện thì nhất thiết B cũng xuất hiện. Ta ký hiệu  $A \subset B$ .

Nếu biến cố A kéo theo biến cố B và ngược lại biến cố B cũng kéo theo biến cố A thì ta có A và B là hai biến cố tương đương. Ta ký hiệu  $A=B$ .

b) Biến cố A và biến cố B được gọi là xung khắc nếu chúng không cùng xuất hiện trong một phép thử.

Trường hợp không xung khắc thì ngược lại.

c) Cho biến cố A thì biến cố không xuất hiện A được gọi là biến cố đối lập với biến cố A. Ta ký hiệu  $\bar{A}$ .

### 1.3.2. Các phép tính giữa các biến cố

a) Tổng của hai biến cố A và B là một biến cố xảy ra khi có ít nhất một trong hai biến cố A, B xảy ra (hoặc A xảy ra hoặc B xảy ra). Ta ký hiệu  $(A+B)$ .

*Ví dụ:* Trong một hộp phấn đựng một số viên phấn đỏ, một số viên phấn xanh và một số viên phấn trắng. Rút hú họa một viên phấn. Khi đó, biến cố rút được viên phấn màu là biến cố tổng của hai biến cố: biến cố rút được viên phấn đỏ và biến cố rút được viên phấn xanh.

b) Tích của hai biến cố A và B là một biến cố xảy ra khi và chỉ khi cả hai biến cố A và B cùng xảy ra. Ta ký hiệu  $(A.B)$ .

*Ví dụ:* Có hai xạ thủ, mỗi người bắn 1 phát và bắn vào cùng một bia. Khi đó biến cố bia bị bắn trúng đúng hai lần là biến cố tích của hai biến cố: biến cố người thứ nhất bắn trúng bia và biến cố người thứ hai bắn trúng bia.

#### 1.4. XÁC SUẤT CỦA TỔNG HAI BIẾN CỐ XUNG KHẮC

Ta thiết lập công thức cho phép tính xác suất của tổng hai biến cố xung khắc theo xác suất của các biến cố thành phần.

**Định lý:** Nếu A, B là hai biến cố xung khắc thì ta có:

$$P(A+B)=P(A)+P(B) \quad (5)$$

Ta chứng minh định lý cho trường hợp các biến cố mà xác suất của chúng thoả mãn định nghĩa cổ điển. Giả sử số kết cục đồng khả năng của phép thử là n, trong đó có  $m_1$  kết cục thích hợp (thuận lợi) cho A và  $m_2$  kết cục thích hợp (thuận lợi) cho B. Theo giả thiết vì A, B xung khắc nên số kết cục thích hợp (thuận lợi) cho biến cố tổng (A+B) sẽ là  $(m_1+m_2)$ . Do đó ta có xác suất tính theo định nghĩa cổ điển:

$$P(A+B) = \frac{m_1+m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B) \quad (6)$$

**Hệ quả:** Nếu A và  $\bar{A}$  là hai biến cố đối lập, thì ta có:

$$P(\bar{A}) = 1 - P(A) \quad (7)$$

Thực vậy, theo giả thiết:  $A + \bar{A} = U$  (biến cố chắc chắn) và  $A \cdot \bar{A} = V$  (biến cố không thể có), nên theo định nghĩa A,  $\bar{A}$  là biến cố xung khắc và ta có thể áp dụng định lý về xác suất của tổng hai biến cố xung khắc cho A và  $\bar{A}$ :

$$P(A + \bar{A}) = P(A) + P(\bar{A}) \quad (8)$$

mặt khác, ta lại có:

$$P(A + \bar{A}) = P(U) = 1 \quad (9)$$

vậy, từ (8) và (9) suy ra:

$$P(A) + P(\bar{A}) = 1$$

hay:

$$P(\bar{A}) = 1 - P(A)$$



Công thức (7) cho phép chuyển việc tính xác suất  $P(A)$  qua việc tính xác suất  $P(\bar{A})$  và trong nhiều trường hợp đưa tới kết quả nhanh hơn. Ngoài ra, công thức (5) có thể mở rộng cho một số hữu hạn các biến cố xung khắc từng đôi một, nghĩa là nếu  $A_1, A_2, \dots, A_n$  là các biến cố xung khắc từng đôi một, thì ta có:

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

*Ví dụ 1:* Trong một hộp phấn có 5 viên phấn xanh, 4 viên phấn đỏ và 11 viên phấn trắng. Rút hù họa một viên phấn. Hãy tính xác suất:

a) Rút ra được viên phấn màu.

b) Rút ra được viên phấn trắng.

*Giải:* a) Tính xác suất rút ra được viên phấn màu. Gọi A là biến cố rút ra được viên phấn đỏ, B là biến cố rút ra được viên phấn xanh. Rõ ràng các biến cố A, B là xung khắc và nếu gọi C là biến cố rút ra được viên phấn màu, thì  $C = (A+B)$ . Áp dụng định lý xác suất của tổng hai biến cố xung khắc và định nghĩa xác suất theo quan điểm cổ điển, ta có:

$$P(C) = P(A + B) = P(A) + P(B) = \frac{4}{20} + \frac{5}{20} = \frac{9}{20}$$

b) Tính xác suất rút ra được viên phấn trắng: Nếu C là biến cố rút ra được viên phấn màu, thì biến cố rút ra được viên phấn trắng rõ ràng là biến cố đối lập của biến cố rút ra được viên phấn màu và sẽ là  $\bar{C}$ . Áp dụng hệ quả về tính xác suất của biến cố đối lập, ta có:

$$P(\bar{C}) = 1 - P(C) = 1 - \frac{9}{20} = \frac{11}{20}$$

*Ví dụ 2:* Một đợt xổ số phát hành N vé, trong đó có M vé có thưởng. Một người mua r vé (với giả thiết  $r < N-M$ ). Tính xác suất để người đó có ít nhất một vé trúng thưởng.

*Giải:* Gọi A là biến cố trong r vé có ít nhất một vé trúng thưởng. Thế thì biến cố đối lập  $\bar{A}$  là biến cố cả r vé đều trượt. Trước tiên ta tính  $P(\bar{A})$ . Vì số kết cục thuận lợi cho  $\bar{A}$  xuất hiện là số tổ hợp  $C_{N-M}^r$ , còn số kết cục đồng khả năng là số tổ hợp  $C_N^r$ , nên ta có:

$$P(\bar{A}) = \frac{C_{N-M}^r}{C_N^r}$$

Từ đó suy ra:

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{C_{N-M}^r}{C_N^r}.$$

## 1.5. XÁC SUẤT CÓ ĐIỀU KIỆN - ĐỊNH LÝ NHÂN XÁC SUẤT

### 1.5.1. Xác suất có điều kiện - Biến cố phụ thuộc và biến cố độc lập

#### a) Bài toán dẫn đến khái niệm

Trong một hộp phấn có 5 viên phấn xanh và 4 viên phấn đỏ. Lấy ra một viên phấn (không hoàn lại), rồi lại lấy ra một viên phấn nữa. Gọi A là biến cố để viên phấn thứ nhất có màu đỏ và B là biến cố để viên phấn thứ hai có màu xanh. Tìm xác suất để viên thứ hai có màu xanh?

*Giải:* Có hai khả năng xảy ra. Khả năng thứ nhất là viên phấn đầu lấy ra là đỏ, tức biến cố A đã xảy ra. Khi đó xác suất để viên thứ hai có màu xanh  $P(B) = \frac{5}{8}$ . Khả năng thứ hai là viên phấn đầu lấy ra là xanh, tức biến cố A đã không xảy ra. Khi đó xác suất để viên thứ hai có màu xanh  $P(B) = \frac{4}{8}$ . Như vậy, ta nói xác suất biến cố B với giả thiết A đã xảy ra là  $5/8$  và xác suất biến cố B với giả thiết A không xảy ra là  $4/8$ .



Từ bài toán trên ta có nhận xét: xác suất  $P(B)$  ở trên là xác suất có điều kiện và hai biến cố A và B là hai biến cố phụ thuộc.

**b) Định nghĩa 1**

Xác suất của biến cố B với giả thiết biến cố A đã xảy ra (ký hiệu là  $P_A(B)$ ) là tỷ số giữa xác suất của tích hai biến cố A, B với xác suất của biến cố A. Ta có:

$$P_A(B) = \frac{P(A.B)}{P(A)} \quad (10)$$

và tương tự ta cũng có:

$$P_B(A) = \frac{P(A.B)}{P(B)} \quad (11)$$

**c) Định nghĩa 2**

Hai biến cố A, B gọi là phụ thuộc nếu sự xảy ra hay không xảy ra của một trong hai biến cố ảnh hưởng đến xác suất của biến cố kia, nghĩa là:

$$P_A(B) \neq P_B(B) \neq P(B) \quad (12)$$

hoặc:

$$P_B(A) \neq P_{\bar{B}}(A) \neq P(A) \quad (13)$$

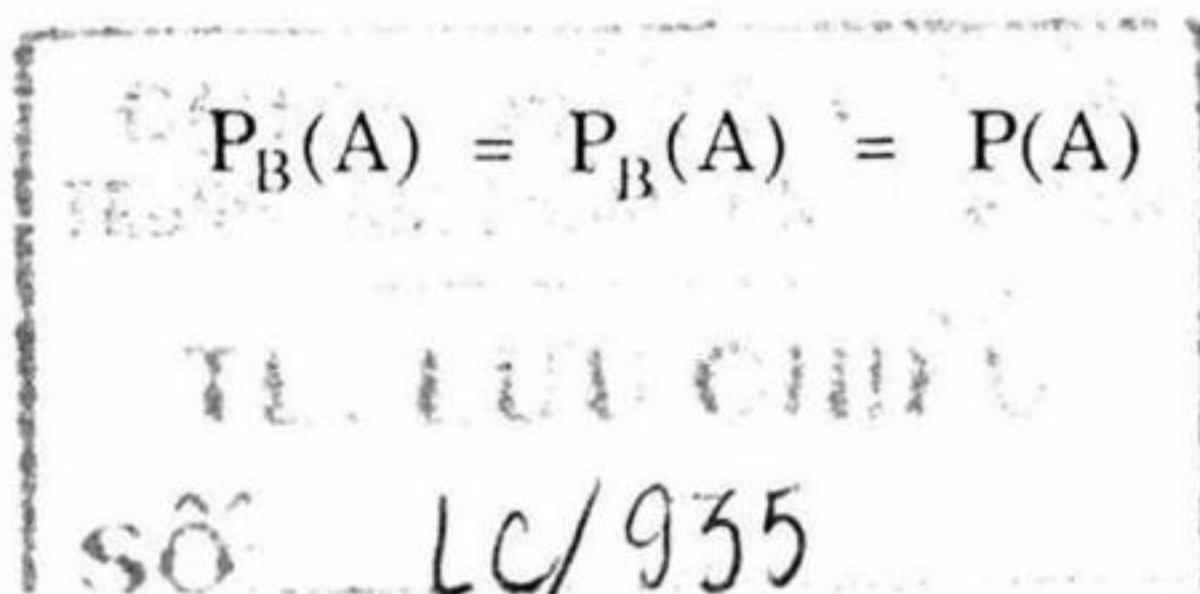
**d) Định nghĩa 3**

Hai biến cố A, B gọi là độc lập với nhau nếu sự xảy ra hay không xảy ra của một trong hai biến cố không ảnh hưởng đến xác suất của biến cố kia, nghĩa là:

$$P_A(B) = P_{\bar{A}}(B) = P(B) \quad (14)$$

hoặc:

$$P_B(A) = P_{\bar{B}}(A) = P(A) \quad (15)$$



### 1.5.2. Định lý nhân xác suất

#### a) Định lý 1

Xác suất của tích hai biến cố phụ thuộc bằng tích của xác suất một trong hai biến cố ấy với xác suất của biến cố còn lại với giả thiết biến cố kia đã xảy ra, nghĩa là:

$$P(A.B) = P(A)P_A(B) = P(B)P_B(A) \quad (16)$$

Công thức (16) suy ra từ công thức (10), (11) trong định nghĩa 1.

#### b) Định lý 2

Xác suất của tích hai biến cố độc lập bằng tích các xác suất của chúng, nghĩa là:

$$P(A.B) = P(A).P(B) \quad (17)$$

Công thức (17) suy ra từ định nghĩa 3 về hai biến cố độc lập (công thức (14), (15)) và định lý nhân xác suất 1 (công thức (16)). Công thức (17) cũng được mở rộng đúng đắn cho số biến cố độc lập lớn hơn 2.

#### c) Ví dụ 1

Một ngăn hộp đựng 3 linh kiện bán dẫn loại I và 7 linh kiện bán dẫn loại II. Một kỹ sư vô tuyến-điện tử lắp máy rút hủ họa một linh kiện (không hoàn lại) và sau đó rút tiếp một linh kiện thứ hai. Hãy tính xác suất để chiếc linh kiện thứ nhất là loại 1 và chiếc linh kiện thứ hai là loại 2.

*Giải:* Gọi A là biến cố chiếc linh kiện thứ nhất thuộc loại I và B là biến cố chiếc linh kiện thứ hai thuộc loại II. Ta cần tính  $P(A.B)$ :

$$P(A.B) = P(A).P_A(B) = \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}$$



#### d) Ví dụ 2

Một thiết bị điện tử gồm 3 bộ phận, cho biết các bộ phận hoạt động độc lập với nhau. Xác suất để trong khoảng thời gian T bộ phận 1, 2, 3 không bị hỏng hóc tương ứng là 0,9; 0,8; 0,7. Tính xác suất có ít nhất một trong ba bộ phận không bị hỏng hóc trong vòng thời gian trên (biến cố H).

*Giải:* Gọi A, B, C tương ứng là biến cố bộ phận I, II, III không bị hỏng hóc trong khoảng thời gian T. Theo giả thiết:

$$P(A) = 0,9 ; P(B) = 0,8 ; P(C) = 0,7$$

Và khi đó xác suất tương ứng để mỗi bộ phận bị hỏng hóc trong khoảng thời gian T tương ứng là:

$$P(\bar{A}) = 1 - P(A) = 1 - 0,9 = 0,1$$

$$P(\bar{B}) = 1 - P(B) = 1 - 0,8 = 0,2$$

$$P(\bar{C}) = 1 - P(C) = 1 - 0,7 = 0,3$$

Xác suất để cả ba bộ phận đều bị hỏng hóc trong khoảng thời gian T là:

$$P(A \cdot B \cdot C) = P(A).P(B).P(C) = 0,1.0,2.0,3 = 0,006$$

Biến cố có ít nhất một bộ phận không bị hỏng hóc là biến cố đối lập của biến cố cả ba bộ phận bị hỏng hóc, do đó ta có:

$$P(H) = 1 - P(\bar{A} \cdot \bar{B} \cdot \bar{C}) = 1 - 0,006 = 0,994$$

## 1.6. XÁC SUẤT CỦA TỔNG HAI BIẾN CỐ KHÔNG XUNG KHẮC

### 1.6.1. Định lý

Xác suất của tổng hai biến cố không xung khắc bằng tổng các xác suất của chúng trừ đi xác suất tích của chúng, nghĩa là:

$$P(A+B)=P(A)+P(B)-P(A.B) \quad (18)$$

*Chứng minh:* Từ khái niệm về mối quan hệ giữa các biến cố và phép tính giữa các biến cố, nếu A, B là hai biến cố không xung khắc, thì ta có thể viết:

$$A+B=A.\overline{B} + \overline{A}.B+A.B \quad (19)$$

Mặt khác, các biến cố  $(A.\overline{B})$ ,  $(\overline{A}.B)$ ,  $(A.B)$  xung khắc với nhau, nên ta có:

$$P(A+B)=P(A.\overline{B})+P(\overline{A}.B)+P(A.B) \quad (20)$$

Cũng suy luận tương tự ta có thể viết:

$$A=A.\overline{B} +A.B$$

$$B=B.\overline{A} +B.A$$

và cũng do biến cố  $(A.\overline{B})$  xung khắc với  $(A.B)$ ; biến cố  $(B.\overline{A})$  xung khắc với  $(B.A)$ , ta lại có:

$$P(A)=P(A.\overline{B})+P(A.B) \quad (21)$$

$$P(B)=P(\overline{A}.B)+P(A.B) \quad (22)$$

Đặt (21), (22) vào (20), ta có:

$$P(A+B)=P(A) - P(A.B) + P(B) - P(A.B) + P(A.B)=P(A) + P(B) - P(A.B)$$

Đó chính là công thức (18) cần chứng minh.

### 1.6.2. Ví dụ

Hai xạ thủ, mỗi người bắn một phát vào bia. Xác suất trúng đích của người thứ nhất là 0,7 ; của người thứ hai là 0,8. Tính xác suất để có ít nhất một phát trúng bia.

*Giải:* Gọi A là biến cố xạ thủ thứ nhất bắn trúng bia; B là biến cố xạ thủ thứ hai bắn trúng bia. Ta cần tìm  $P(A+B)$ . Theo điều kiện của bài toán, ta có  $P(A)=0,7$ ;  $P(B)=0,8$  và hai biến cố A, B là hai biến cố không xung khắc, độc lập với nhau. Suy ra:

$$P(A+B) = P(A) + P(B) - P(A.B)$$

$$P(A.B) = P(A).P(B)$$

$$\Rightarrow P(A+B) = P(A) + P(B) - P(A).P(B) = 0,7+0,8 - 0,7.0,8 = 0,94$$



## 1.7. ĐỘ TIN CẬY CỦA MỘT HỆ THỐNG THIẾT BỊ

Để hiểu rõ hơn cách vận dụng các qui tắc tính xác suất đã trình bày ở trên ta đưa thêm một ứng dụng thực tế như sau:

Giả sử có một hệ thống thiết bị gồm nhiều linh kiện ghép thành. Ta gọi xác suất để một linh kiện hoạt động tốt (không có sự cố) trong khoảng thời gian  $T$  (1 giờ, 24 giờ hay một đơn vị thời gian nào đó) là độ tin cậy của linh kiện ấy. Tương tự ta gọi độ tin cậy của một hệ thống là xác suất để hệ hoạt động tốt trong khoảng thời gian ấn định.

Một vấn đề kỹ thuật đặt ra là: cho biết độ tin cậy của từng linh kiện, hãy tính độ tin cậy của hệ thống.

Lời giải bài toán tùy thuộc vào cách ghép nối các linh kiện:

- Nếu tất cả các linh kiện đều mắc nối tiếp (hình 3a) thì điều kiện để hệ thống hoạt động tốt là mọi linh kiện đều hoạt động tốt. Gọi  $p_i$  là độ tin cậy của linh kiện  $\alpha_i$  ( $i=1, 2, \dots, n$ ). Giả sử các linh kiện hoạt động độc lập lẫn nhau, khi đó theo qui tắc nhân xác suất ta có độ tin cậy của hệ thống là:

$$P(H) = p_1 p_2 \dots p_n \quad (23)$$

Trường hợp riêng khi các linh kiện  $\alpha_1 \dots \alpha_n$  cùng loại thì  $p_1 = \dots = p_n = p$ , do đó  $P(H) = p^n$ .

Vì  $0 < p < 1$  (trên thực tế) nên nếu  $n$  lớn thì  $P(H)$  sẽ bé, nghĩa là khi ghép nối tiếp nhiều linh kiện thì độ tin cậy của hệ thống giảm xuống rõ rệt. Chẳng hạn, nếu mỗi linh kiện có độ tin cậy khá cao ( $p=0,99$ ) thì khi ghép nối tiếp 1000 chiếc, độ tin cậy của toàn mạch giảm xuống chỉ còn xấp xỉ 0,00004.

- Nếu  $n$  linh kiện mắc song song thành  $n$  nhánh, mỗi nhánh 1 linh kiện thì điều kiện để hệ thống hoạt động tốt là có ít nhất một linh kiện hoạt động tốt (hình 3b).

Để tính độ tin cậy  $P(H)$  của hệ thống ta dùng qui tắc tính xác suất  $P(\bar{H})$  của sự kiện đối lập ( $\bar{H}$  : sự kiện tất cả  $n$  nhánh ngừng hoạt động): Ta có:

$$P(\bar{H}) = (1 - p_1)(1 - p_2) \dots (1 - p_n)$$

Do đó:

$$P(H) = 1 - P(\bar{H}) = 1 - (1 - p_1)(1 - p_2) \dots (1 - p_n) \quad (24)$$

Trường hợp riêng khi các linh kiện cùng loại, ta có:

$$p_1 = p_2 = \dots = p_n = p \text{ và } P(H) = 1 - (1 - p)^n = 1 - q^n \quad (25)$$

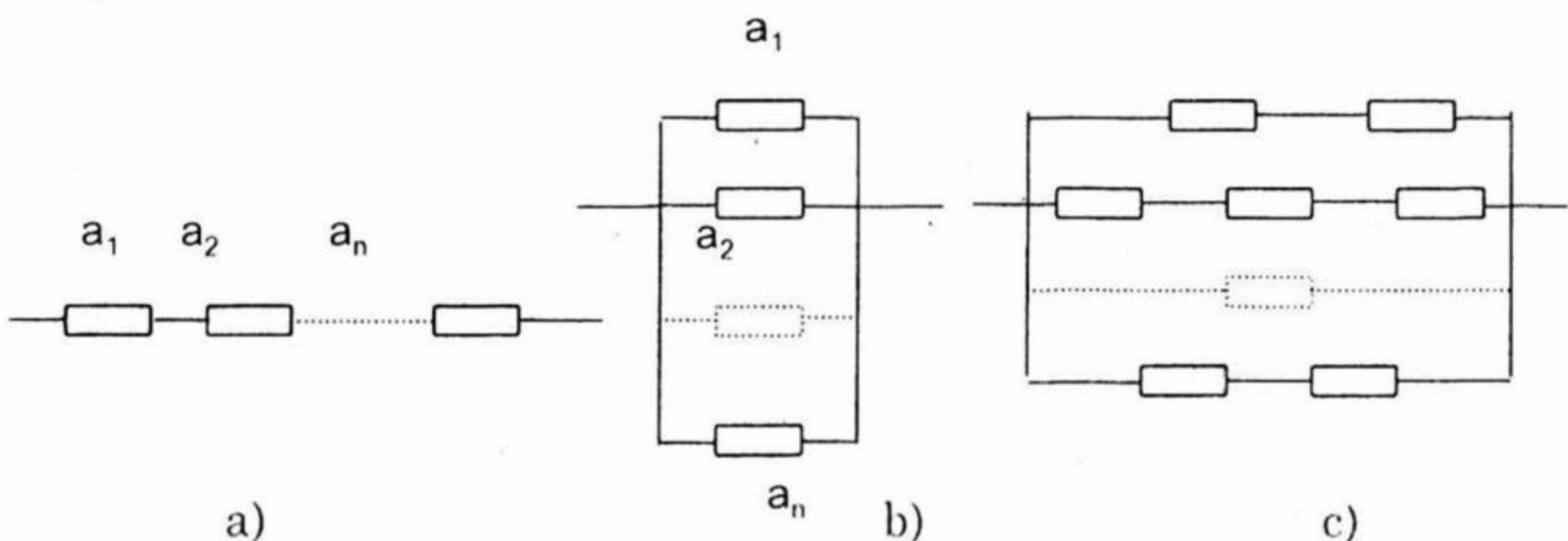
trong đó  $q=1-p$ .

Vì trên thực tế  $0 < q < 1$  nên nếu  $n$  khá lớn thì  $P(H)$  sẽ rất gần 1, thành thử khi ghép song song nhiều linh kiện, độ tin cậy của hệ thống sẽ tăng lên rõ rệt.

- Nếu hệ thống ghép hỗn hợp (hình 3c) thì trước tiên ta dùng công thức (23) để tính độ tin cậy của mỗi nhánh, sau đó coi mỗi nhánh như một linh kiện với độ tin cậy vừa tìm được rồi áp dụng công thức (24) để tính  $P(H)$ .

Chẳng hạn có tất cả  $m.n$  linh kiện cùng loại, mỗi chiếc có độ tin cậy  $p$ , ghép thành  $n$  nhánh, mỗi nhánh gồm  $m$  linh kiện thì độ tin cậy của toàn hệ thống sẽ là:

$$P(H) = 1 - (1 - p^m)^n$$



**Hình 3**



*Hệ thống dự phòng.* Để tăng độ tin cậy của hệ thống người ta thường ghép song song với hệ thống chính một hoặc nhiều hệ thống dự phòng. Có hai cách ghép hệ thống dự phòng: dự phòng toàn bộ và dự phòng từng cụm.

Giả sử có một hệ thống gồm  $n$  linh kiện giống nhau đấu nối tiếp. Độ tin cậy của hệ đã cho là  $p^n$ , trong đó  $p$  là độ tin cậy của mỗi linh kiện.

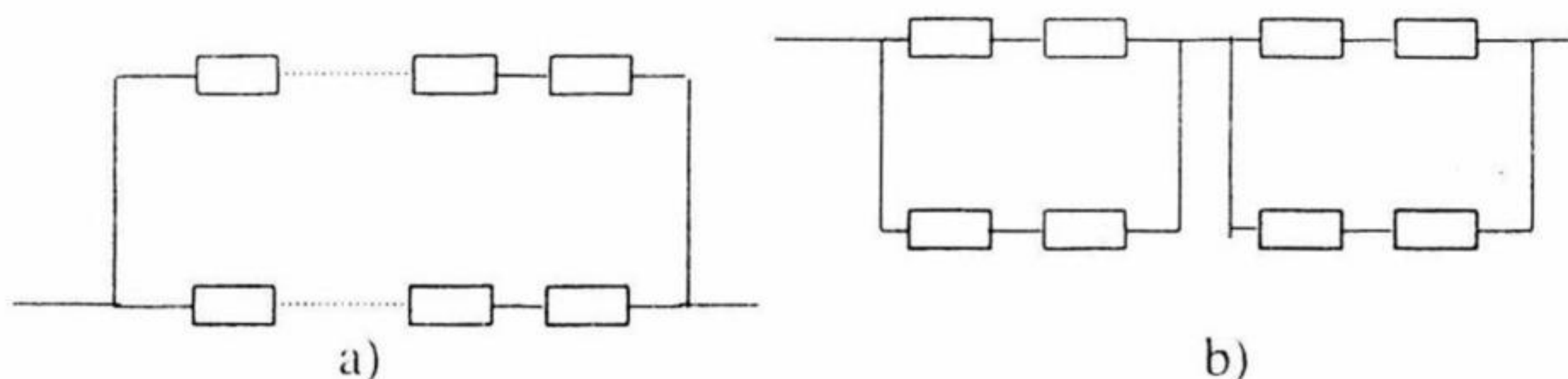
Nếu ghép dự phòng thêm một dãy như thế (hình 4a) thì độ tin cậy toàn bộ hệ mới sẽ là:

$$P(H) = 1 - (1 - p^n)^2 = 2p^n - p^{2n}$$

Khi  $n$  khá lớn  $P(H) \approx 2p^n$ , nghĩa là độ tin cậy tăng gần gấp đôi.

Cũng có thể dự phòng theo  $r$  cụm, mỗi cụm gồm  $s$  linh kiện ( $n=r.s$ ) (hình 4b). Độ tin cậy của toàn hệ mới sẽ là:

$$P(H) = [1 - (1 - p^s)^2]^r = (2p^s - p^{2s})^r$$



**Hình 4**

Khi  $n$  khá lớn và  $s$  khá lớn ta có:

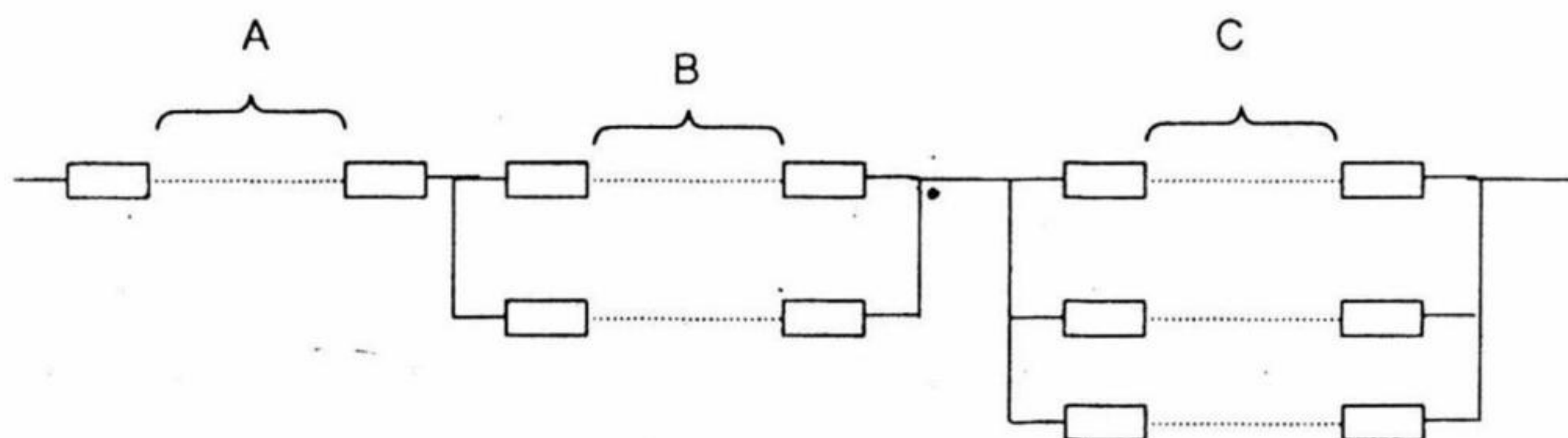
$$P(H) \approx (2p^s)^r = 2^r p^n$$

nghĩa là độ tin cậy tăng gần gấp  $2^r$  lần.

Thông thường với các yêu cầu qui định của kỹ thuật và tùy theo đặc tính của từng loại linh kiện, ta có thể bố trí những cụm dự phòng với số dãy song song trong từng cụm khác nhau để làm tăng độ tin cậy của hệ thống sao cho giá thành của thiết bị không cao và việc chế tạo không quá phức tạp.

*Ví dụ:* Một hệ thống gồm 40 linh kiện loại A với độ tin cậy mỗi chiếc  $p_A=0,99$ ; 25 linh kiện loại B với độ tin cậy mỗi chiếc  $p_B=0,9$  và 5 linh kiện loại C với độ tin cậy mỗi chiếc  $p_C=0,75$ . Giá thành mỗi linh kiện loại A, B, C tương ứng là 1, 1, 5 (đơn vị tiền).

Hãy lập một hệ thống dự phòng toàn bộ, đánh giá độ tin cậy và giá thành rồi so sánh với một hệ thống dự phòng từng cụm theo kiểu không dùng loại A, lắp thêm một bộ loại B và hai bộ loại C (hình 5).



**Hình 5**

*Giải:* Theo kiểu mắc hệ thống dự phòng toàn bộ thì độ tin cậy của toàn hệ mới sẽ là  $P_1=1 - (1 - p)^2$ , trong đó  $p$  là độ tin cậy của hệ thống ban đầu:

$$p = (0,99)^{40} \cdot (0,9)^{25} \cdot (0,75)^5.$$

Vì  $p$  khá bé nên  $P_1 \approx 2p$  (gấp đôi).

Giá thành của hệ thống dự phòng sẽ là:

$$40 \cdot 1 + 25 \cdot 1 + 5 \cdot 5 = 90 \text{ đơn vị}$$

Bây giờ ta đánh giá độ tin cậy theo kiểu thứ hai. Muốn thế đầu tiên ta tìm độ tin cậy của mỗi cụm, gọi  $P_A$ ,  $P_B$ ,  $P_C$  tương ứng là độ tin cậy của cụm loại A, loại B và loại C. Ta có:

$$P_A = (p_A)^{40} = (0,99)^{40};$$

$$P_B = 1 - (1 - p_B^{25})^2 = 2p_B^{25} - p_B^{50} \approx 2p_B^{25} = 2 \cdot (0,9)^{25}$$

$$P_C = 1 - (1 - p_C^5)^3 \approx 3p_C^5 = 3 \cdot (0,75)^5$$

Độ tin cậy của toàn hệ thống mới sẽ là:



$$P_2 = P_A P_B P_C \approx (0,99)^{10} \cdot 2(0,90)^{25} 3(0,75)^5 = 6p$$

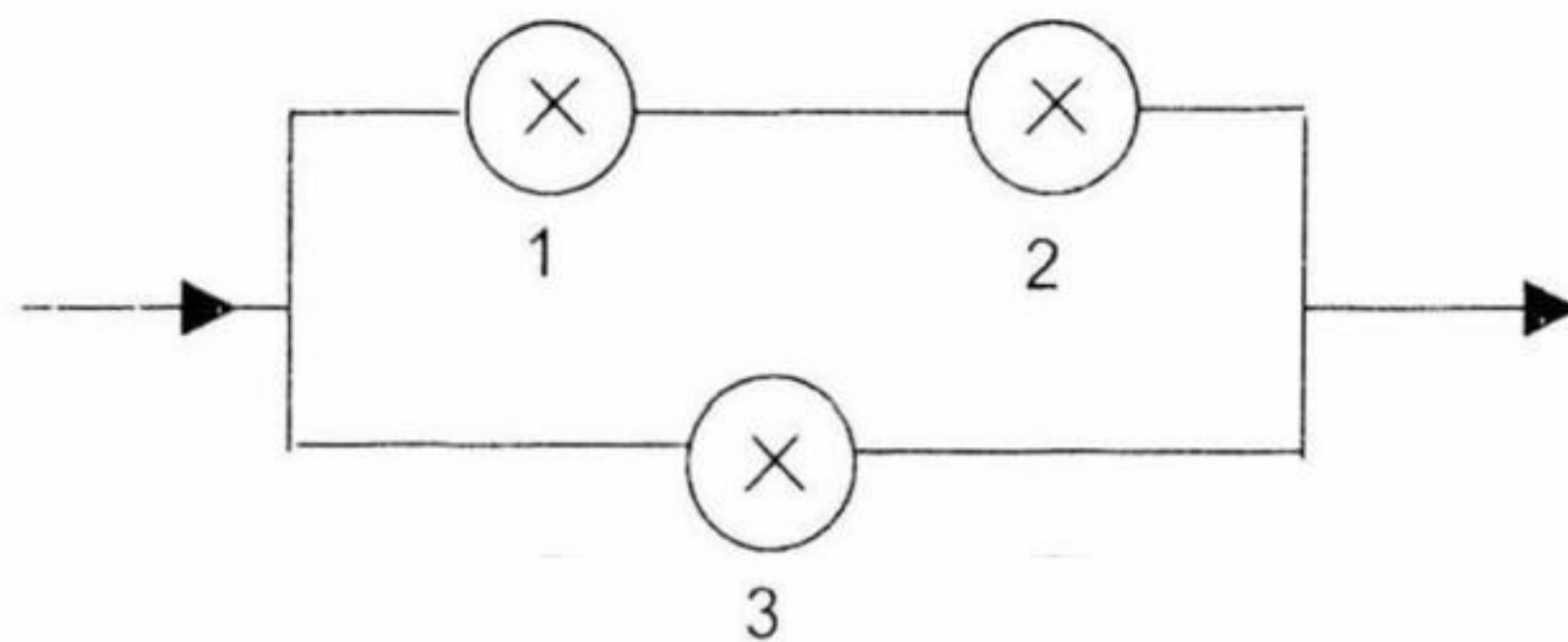
So với  $P_1$  ta thấy  $P_2 \approx 3P_1$ , tức độ tin cậy tăng gần gấp 3 lần so với kiểu thứ nhất mà giá thành lại hạ hơn vì theo kiểu này hệ dự phòng chỉ tốn có  $25 \cdot 1 + 10 \cdot 5 = 75$  đơn vị tiền.

## BÀI TẬP CHƯƠNG 1

**Bài 1:** Khi nào thì có đẳng thức  $A + B = A$  ?

**Giải:** Đẳng thức  $A + B = A$  sẽ đúng nếu B kéo theo A (hay việc xảy ra B luôn kéo theo xảy ra A).

**Bài 2:** Cho sơ đồ mạng điện trên hình 6, gồm 3 bóng đèn. Việc mạng mất điện (sự kiện A) chỉ có thể xảy ra do cháy các bóng đèn (ký hiệu là các sự kiện  $A_1, A_2$  và  $A_3$ ). Hãy biểu diễn A theo các  $A_i, i=1, 2, 3$ .



**Hình 6**

**Giải:** Biến cố A xảy ra khi xảy ra một trong ba trường hợp:

- a) Cả ba bóng đều cháy;
- b) Cháy hai bóng 1 và 3;
- c) Cháy hai bóng 2 và 3.

Từ đó ta có:

$$A = A_1 A_2 A_3 + A_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3.$$

Có thể dùng tính chất của việc mắc song song và nối tiếp các bóng để có một biểu diễn khác như sau:

$$A = (A_1 + A_2) A_3.$$

**Bài 3:** Biến cố A- có ít nhất một trong 4 sản phẩm là phế phẩm;  
B- số phế phẩm không ít hơn 2. Các biến cố sau là gì:

- a)  $\bar{A}$  ;
- b)  $\bar{B}$  ;
- c)  $\bar{A} B$  ?

**Giải:** a) Dễ thấy  $\bar{A}$  - không có phế phẩm hay cả 4 sản phẩm đều tốt;

b)  $\bar{B}$  - hoặc có một phế phẩm hoặc không có phế phẩm; hay  $\bar{B}$  - có nhiều nhất một phế phẩm (hoặc có ít nhất ba chính phẩm);

c)  $A \bar{B}$  - có đúng 1 phế phẩm;

d)  $\bar{A} B = V$  (không xảy ra).

**Bài 4:** Trong 52 quân bài tú lơ khơ, bỏ 20 quân từ bộ 2 đến bộ 6, rồi chia đều cho 3 người, mỗi người 10 quân, còn 2 quân giữ lại để rút (làm “nọc”). Tính xác suất để 2 quân còn giữ lại để rút là quân “át”.

**Giải:** Gọi A là biến cố 2 quân bài còn giữ lại để rút là quân “át”. Thế thì số kết cục thuận lợi cho A xuất hiện là số tổ hợp  $C_1^2$ , số kết cục đồng khả năng xảy ra là số tổ hợp  $C_{32}^2$ . Ta có:

$$P(A) = \frac{C_1^2}{C_{32}^2} = \frac{6}{496}.$$

**Bài 5:** Có 5 mảnh bìa được đánh số từ 1 đến 5. Chọn hú họa liên tiếp ra 3 mảnh và xếp thành một số có 3 chữ số. Tìm xác suất để số đó là số chẵn.

**Giải:** Do ta chọn liên tiếp 3 mảnh không hoàn lại và có để ý đến thứ tự nên số cách chọn sẽ chính là số các chỉnh hợp chập 3 đến 5:

$$A_5^3 = 5 \cdot 4 \cdot 3 = 60.$$

Để có số chẵn thì số lấy cuối cùng phải là chẵn, tức là hoặc là 2, hoặc là 4. Ứng với mỗi số đó ta có số cách lấy ra 2 số trước nó là  $4 \cdot 3 = 12$ . Từ đó số cách chọn ra được số chẵn theo yêu cầu là  $12 \cdot 2 = 24$ . Vậy xác suất cần tìm là



$$\frac{24}{60} = 0.4.$$

**Bài 6:** Trong hộp bi có 6 viên đỏ và 4 viên trắng cùng kích cỡ. Rút hù họa ra 2 viên bi. Tính xác suất để trong đó có:

- a) 2 viên đỏ;
- b) Ít nhất 1 viên đỏ;
- c) Viên thứ 2 màu đỏ.

**Giải:** Nếu ta quan tâm đến thứ tự của hai viên bi, số cách rút ra 2 bi sẽ là  $A_{10}^2 = 10 \cdot 9 = 90$ .

a) Số cách thuận lợi để rút ra 2 bi đỏ trong trường hợp này là  $6 \cdot 5 = 30$ . Từ đó xác suất cần tìm là  $30/90 = 1/3$ . Có thể dùng khái niệm tổ hợp để tính xác suất: tổng số cách rút ra 2 bi không để ý đến thứ tự là  $C_{10}^2$ , còn số cách thuận lợi rút được 2 bi đỏ là  $C_6^2$ , từ đó tìm ra cùng kết quả như trên.

b) Cách tính trực tiếp phải dùng lại kết quả trên. Ta có thể dễ dàng tính xác suất xảy ra biến cố đối lập “không có bi đỏ” hay “cả 2 bi trắng” là  $C_4^2 / C_{10}^2 = 2/5$ . Từ đó xác suất cần tìm là  $1 - 2/5 = 3/5$ .

c) Gọi A là biến cố bi thứ hai màu đỏ. Số cách thuận lợi cho A bao gồm: 6.5 cách đối với trường hợp viên đầu cũng đỏ và 4.6 cách đối với trường hợp viên đầu trắng; từ đó xác suất  $P(A) = (30 + 24)/90 = 3/5$ .

Chú ý nếu theo cách không quan tâm đến thứ tự thì mọi việc đơn giản hơn: A sẽ tương đương với biến cố viên đầu là đỏ, xác suất để rút được một viên bi đỏ rất dễ tính là  $6/10 = 3/5$ .

**Bài 7:** Tìm xác suất để khi xếp ngẫu nhiên 5 người quanh 1 chiếc bàn tròn 5 ghế thì hai người định trước được ngồi cạnh nhau.

**Giải:** Dễ thấy tổng số cách xếp 5 người là số các hoán vị của 5 và bằng  $P_5 = 5! = 120$ . Do vai trò của 5 người như nhau nên không mất tính tổng quát ta có thể bắt đầu tính từ bất kỳ người nào, chẳng hạn từ một trong hai người định trước. Người thứ nhất trong hai người đó chỉ



có 5 cách xếp và để được ngồi cạnh người đó người thứ hai chỉ còn 2 cách xếp; còn đối với 3 người còn lại có tất cả  $3!$  cách xếp. Tóm lại số cách thuận lợi cho biến cố bài ra sẽ là  $5.2.3!=60$ ; từ đó xác suất cần tìm là  $60/120=1/2$ .

**Bài 8:** Trong một buổi liên hoan có 6 cặp nam nữ, trong đó có 3 cặp là vợ chồng. Chọn hú họa ra 3 người. Tìm các xác suất để trong đó:

- a) Có đúng 1 nam;
- b) Cả 3 đều là nữ;
- c) Không có cặp vợ chồng nào.

**Giải:** Tổng số kết cục của phép thử chọn hú họa ra ba người chính là  $C_{12}^3 = 220$ .

a) Để chọn được đúng 1 nam (và có nghĩa là 2 người còn lại là nữ) sẽ có  $C_6^1 \cdot C_6^2 = 90$  cách. Từ đó xác suất cần tìm là  $90/220=9/22$ .

b) Tương tự số cách chọn được 3 nữ là  $C_6^3 = 20$  và xác suất cần tìm là  $1/11$ .

c) Việc tìm trực tiếp số cách thuận lợi cho A - biến cố không có cặp vợ chồng nào trong số 3 người khá phức tạp. Ta tính xác suất của  $\bar{A}$  là sự kiện có ít nhất 1 cặp vợ chồng, và vì chỉ có 3 người nên đó chính là biến cố có đúng 1 cặp vợ chồng. Cặp vợ chồng đó có 3 cách chọn, còn người thứ ba sẽ có 10 cách chọn trong số 10 người còn lại, từ đó:

$$P(A)=1- P(\bar{A})=1- 3.10/220=19/22.$$

**Bài 9:** Có 10 mảnh bìa được đánh số từ 0 đến 9. Lấy hú họa ra hai mảnh bìa và xếp thành một số có 2 chữ số. Tìm xác suất để số đó chia hết cho 18.

**Giải:** Ký hiệu số xếp được là  $N=10a + b$ , trong đó  $0 \leq a, b \leq 9$ . Ta thấy N chia hết cho 18 thì phải chẵn (b chẵn) và chia hết cho 9 ( $a+b$  chia hết cho 9). Dễ thấy tổng số cách chọn ra 2 mảnh bìa và xếp thành một số là  $10.9 = 90$ . Số cách chọn thuận lợi cho số N chia hết cho 18 là 5 (đó là các số 18, 36, 54, 72 và 90). Từ đó xác suất cần tìm là  $5/90=1/18$ .



**Bài 10:** Một lô hàng có  $n$  sản phẩm với  $m$  phế phẩm. Tìm xác suất để khi chọn hủ hóa ra  $k$  sản phẩm thì có đúng  $l$  phế phẩm.

**Giải:** Số cách chọn ngẫu nhiên ra  $k$  sản phẩm từ  $n$  sẽ là  $C_n^k$ . Số cách chọn thuận lợi cho biến cố cần tìm xác suất chính là tích của số cách chọn  $l$  phế phẩm từ  $m$  với số cách chọn  $k-l$  chính phẩm từ  $n-m$  sản phẩm tốt. Xác suất cần tìm dễ dàng tính được bằng:

$$\frac{C_m^l C_{n-m}^{k-l}}{C_n^k}.$$

**Bài 11:** Xếp ngẫu nhiên 10 khách đi tàu lên 3 toa tàu hoả. Hãy tìm các xác suất:

- Toa đầu có 3 khách;
- Toa đầu có 3 khách và toa thứ hai có 4 khách;
- Một toa có 3 khách và một toa khác có 4 khách (toa còn lại tất nhiên có 3 khách trong cả hai trường hợp b và c).

**Giải:** Mỗi khách có 3 khả năng khác nhau lên các toa tàu, vậy 10 người sẽ có  $3^{10}$  cách lên tàu khác nhau (chính là số chỉnh hợp lặp chập 10 từ 3 phần tử  $\tilde{A}_3^{10} = 3^{10}$ ).

a) Để toa đầu có 3 khách sẽ có  $C_{10}^3$  cách xếp; sau đó 7 khách còn lại sẽ có  $2^7$  cách xếp lên hai toa còn lại. Từ đó số cách xếp thuận lợi sẽ là  $C_{10}^3 \cdot 2^7$  và xác suất cần tìm là:

$$\frac{C_{10}^3 \cdot 2^7}{3^{10}} = \frac{5120}{19683}.$$

b) Để toa đầu có 3 khách sẽ có  $C_{10}^3$  cách xếp, sau đó để toa hai có 4 khách có  $C_7^4$  cách và để toa ba có 3 khách còn lại là  $C_3^3$ . Xác suất cần tìm sẽ là:

$$\frac{C_{10}^3 \cdot C_7^4 \cdot C_3^3}{3^{10}} = \frac{1400}{19683}$$

c) Dễ thấy số cách thuận lợi cho trường hợp này chính bằng số cách thuận lợi cho trường hợp b) nhân 3. Từ đó xác suất cần tìm là:

$$\frac{3 \cdot C_{10}^3 \cdot C_7^4 \cdot C_3^3}{3^{10}} = \frac{2800}{6561}.$$

**Bài 12:** Một lớp học có 30 sinh viên trong đó có 5 giỏi, 10 khá và 10 trung bình, còn lại là học sinh yếu. Chọn hù họa ra 3 người, hãy tìm các xác suất:

- a) Cả 3 đều là học sinh yếu;
- b) Có ít nhất một học sinh giỏi.

**Giải:** Số cách chọn ra 3 người trong số 30 người dễ thấy là  $C_{30}^3 = 4060$ .

a) Để cả ba đều yếu, có nghĩa là phải chọn được 3 người trong số 5 học sinh yếu. Số cách chọn đó là  $C_5^3 = 10$  và xác suất cần tìm sẽ là:

$$\frac{10}{4060} = \frac{1}{406}$$

b) Ta tính xác suất của biến cố đối lập: không có học sinh giỏi trong số 3 người được chọn ngẫu nhiên đó. Số cách thuận lợi chính là số nhóm gồm 3 phần tử từ 25 phần tử (số học sinh không phải là học sinh giỏi là 25). Từ đó dễ dàng thấy xác suất cần tìm là:

$$1 - \frac{C_{25}^3}{C_{30}^3} = 1 - \frac{115}{203} = \frac{88}{203}.$$

**Bài 13:** Chia thành hai phần bằng nhau 10 viên bi, trong đó có 4 bi đỏ và 6 bi xanh. Tìm xác suất để mỗi phần đều cùng số bi đỏ và bi xanh.

**Giải:** Số cách chia thành hai phần có số bi như nhau chính là số cách chọn ra 5 viên bi từ 10 và đó là  $C_{10}^5$ . Để một phần có số viên bi đỏ và xanh giống phần kia sẽ có  $C_4^2 \cdot C_6^3$  cách chia thuận lợi. Từ đó xác suất cần tìm là:

$$\frac{C_4^2 C_6^3}{C_{10}^5} = \frac{10}{21}$$

**Bài 14:** Đường dây cáp ngầm nối một tổng đài với một trạm dài 1km.

Tính xác suất của sự kiện dây cáp bị đứt tại nơi cách tổng đài không dưới 800m.



**Giải:** Rõ ràng nếu dây cáp đồng chất thì khả năng nó bị đứt tại một điểm bất kỳ là như nhau và tập các kết cục đồng khả năng có thể biểu thị bằng đoạn nối tổng dài với trạm. Các kết cục thuận lợi cho biến cố A- chỗ đứt cách tổng dài không dưới 800m được biểu thị bởi đoạn có độ dài 200m. Từ đó:

$$P(A) = \frac{200}{1000} = \frac{1}{5}$$

Bạn đọc thử tìm xác suất của biến cố trên với giả thiết là càng cách xa tổng dài khả năng dây cáp bị đứt càng lớn (tức là tỉ lệ thuận với khoảng cách từ điểm đứt tới tổng dài).

**Bài 15:** Cho một đoạn thẳng và bẻ gãy ngẫu nhiên thành 3 đoạn. Tìm xác suất để 3 đoạn đó tạo thành được 1 tam giác.

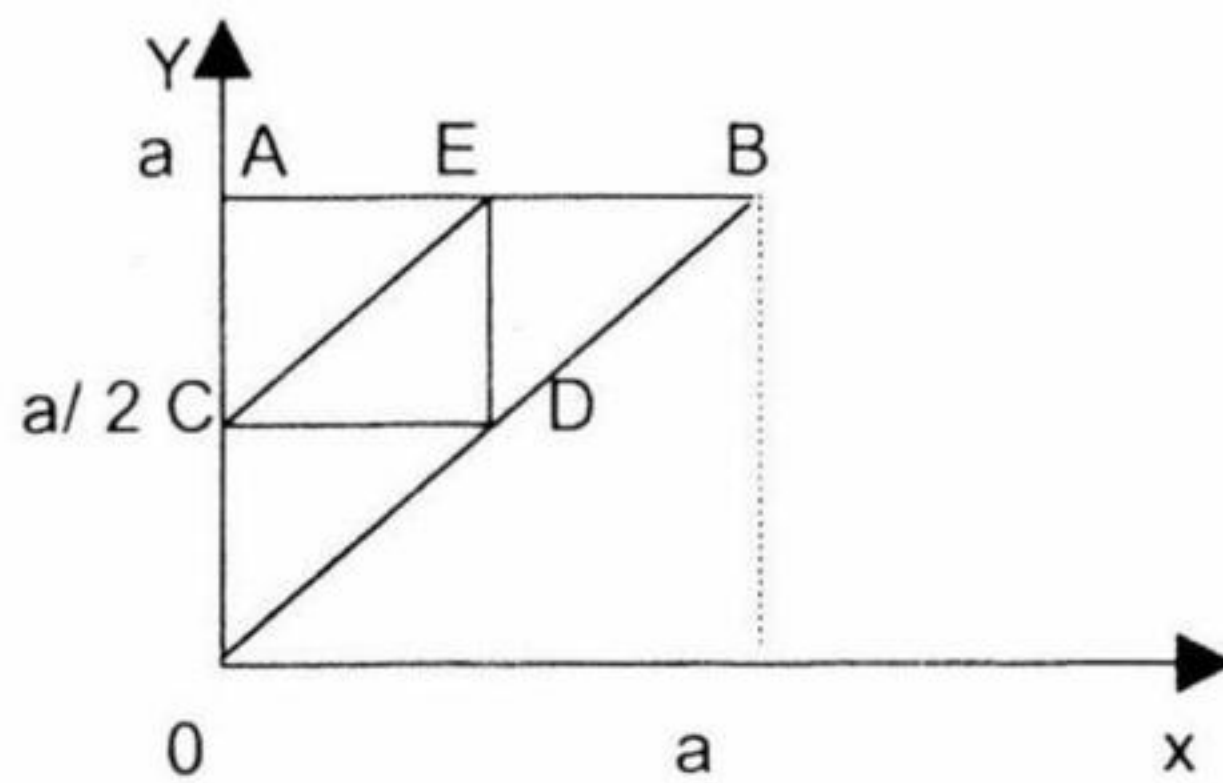
**Giải:** Như bài 13 ta dùng định nghĩa theo hình học. Coi đoạn thẳng là 1 đoạn trên trục số từ 0 đến a. Ký hiệu x là tọa độ điểm chia thứ nhất và y là tọa độ điểm chia thứ hai thì dễ thấy  $0 < x < y < a$  và ba đoạn sẽ có các độ dài tương ứng x, y-x và a-y.

Đặt tương ứng mỗi cách chia với 1 điểm trong hệ tọa độ Để các M(x,y), miền đồng khả năng tương ứng sẽ là tam giác AOB (xem hình 7). Ta cần tìm miền thuận lợi cho sự kiện đầu bài yêu cầu. Muốn tạo ra một tam giác thì tổng 2 cạnh phải lớn hơn cạnh còn lại:

$$x + (y-x) > a - y \text{ nên } y > \frac{a}{2}$$

$$x + (a-y) > y-x \text{ nên } y < x + \frac{a}{2}$$

$$-x + (a-y) > x \text{ nên } x < \frac{a}{2}$$



**Hình 7**

Từ đó miền thuận lợi cần tìm là tam giác CDE. Do diện tích tam giác này bằng  $1/4$  diện tích tam giác AOB nên xác suất cần tìm là

$$\frac{S_{CDE}}{S_{AOB}} = \frac{1}{4}.$$

**Bài 16:** Một tổ có 4 nam và 3 nữ. Chọn liên tiếp ra hai người. Tìm xác suất để:

- Cả hai là nữ;
- Có một nam, một nữ.

**Giải:** Đặt  $A_i$  là biến cố chọn được nữ ở lần  $i$ , và  $B_i$  là biến cố chọn được nam ở lần  $i$ ,  $i=1,2$ .

- Gọi  $A$  là biến cố chọn được 2 nữ, dễ thấy  $A = A_1 \cdot A_2$  và ta có:

$$P(A) = P(A_1 A_2) = P(A_1)P_{A_1}(A_2) = \frac{3}{7} \cdot \frac{2}{6} = \frac{1}{7}.$$

- Gọi  $B$  là biến cố chọn được một nam, một nữ. Có thể chứng tỏ  $B = A_1 B_2 + B_1 A_2$ . Do các số hạng của tổng xung khắc:

$$\begin{aligned} P(B) &= P(A_1 B_2) + P(B_1 A_2) = P(A_1)P_{A_1}(B_2) + P(B_1)P_{B_1}(A_2) = \\ &= \frac{3}{7} \cdot \frac{4}{6} + \frac{4}{7} \cdot \frac{3}{6} = \frac{4}{7}. \end{aligned}$$

**Bài 17:** Có hai hộp bút: hộp 1 có 2 bút đỏ và 10 xanh; hộp 2 có 8 đỏ và 4 xanh. Chọn ngẫu nhiên từ mỗi hộp ra một bút, tìm xác suất để có 1 bút xanh, 1 bút đỏ.



**Giải:** Ta gọi  $A_i$  là biến cố chọn được bút đỏ từ hộp thứ  $i$  và  $B_i$  là biến cố chọn được bút xanh từ hộp thứ  $i$  ( $i=1,2$ ). Theo đầu bài để khi rút hũ họa ra 2 bút có một xanh, một đỏ ta có 2 trường hợp: hoặc chọn bút xanh từ hộp 1, bút đỏ từ hộp 2, hoặc ngược lại, từ đó xác suất cần tìm:

$$\begin{aligned} P(A_1B_2 + A_2B_1) &= P(A_1B_2) + P(A_2B_1) \\ &= P(A_1)P(B_2) + P(A_2)P(B_1) \\ &= \frac{2}{12} \cdot \frac{4}{12} + \frac{8}{12} \cdot \frac{10}{12} = \frac{11}{18} \end{aligned}$$

(để ý là  $A_i$  độc lập với  $B_j$ ,  $j \neq i$ , và  $A_1 = \bar{B}_1$ ;  $A_2 = \bar{B}_2$ ).

**Bài 18:** Một phòng điều trị có 3 bệnh nhân bệnh nặng với xác suất cần cấp cứu trong vòng một giờ của các bệnh nhân tương ứng là 0,7; 0,8 và 0,9. Tìm các xác suất sao cho trong vòng một giờ:

- Có hai bệnh nhân cần cấp cứu;
- Có ít nhất một bệnh nhân không cần cấp cứu.

**Giải:** Đặt  $A_i$  là biến cố có 2 bệnh nhân cần cấp cứu, dễ thấy có thể xảy ra 3 trường hợp khác nhau và:

$$A = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3.$$

Do tính xung khắc của các số hạng và tính độc lập của các  $A_i$  (và  $\bar{A}_i$ ) nên:

$$\begin{aligned} P(A) &= P(A_1A_2\bar{A}_3) + P(A_1\bar{A}_2A_3) + P(\bar{A}_1A_2A_3) \\ &= P(A_1)P(A_2)P(\bar{A}_3) + P(A_1)P(\bar{A}_2)P(A_3) + P(\bar{A}_1)P(A_2)P(A_3) \\ &= 0,7 \cdot 0,8 \cdot 0,1 + 0,7 \cdot 0,2 \cdot 0,9 + 0,3 \cdot 0,8 \cdot 0,9 = 0,398 \end{aligned}$$

b) Gọi  $B$  là biến cố có ít nhất 1 bệnh nhân không cần cấp cứu, dễ thấy  $\bar{B}$  là biến cố không có bệnh nhân không cần cấp cứu tức là tất cả đều cần cấp cứu. Rõ ràng việc tính  $P(\bar{B})$  dễ dàng hơn nhiều so với tính  $P(B)$ , từ đó:

$$\begin{aligned} P(B) &= 1 - P(\bar{B}) = 1 - P(A_1A_2A_3) \\ &= 1 - 0,7 \cdot 0,8 \cdot 0,9 = 0,496. \end{aligned}$$

**Bài 19:** Biết xác suất để một học sinh thi đạt yêu cầu ở lần thứ  $i$  là  $p_i$  ( $i=1,2$ ). Tìm xác suất để học sinh đó đạt yêu cầu trong kỳ thi biết rằng mỗi học sinh được phép thi tối đa 2 lần.

**Giải:** Gọi  $A_i$  là biến cố học sinh đó thi đạt ở lần thi thứ  $i$ , còn  $A$  là biến cố thi đạt yêu cầu của kỳ thi. Dễ dàng thấy  $A=A_1+\bar{A}_1 A_2$ .

Dùng các công thức cộng và nhân xác suất có để ý đến tính độc lập và xung khắc, ta có:

$$\begin{aligned} P(A) &= P(A_1 + \bar{A}_1 A_2) = P(A_1) + P(\bar{A}_1 A_2) = P(A_1) + P(\bar{A}_1)P(A_2) = \\ &= p_1 + (1-p_1)p_2. \end{aligned}$$

**Bài 20:** Một lô hàng gồm 100 sản phẩm, trong đó có 5 phế phẩm. Lô hàng được chấp nhận nếu chọn hú họa ra 50 sản phẩm để kiểm tra thì số phế phẩm không quá 1. Tìm xác suất để lô hàng được chấp nhận.

**Giải:** Gọi  $A$  là biến cố lô hàng được chấp nhận,  $A_i$  - trong số 50 sản phẩm chọn ra có  $i$  phế phẩm ( $i=0,1$ ), từ đó  $A=A_0+A_1$ . Do  $A_0$  và  $A_1$  xung khắc nên:

$$P(A) = P(A_0) + P(A_1) = \frac{C_{95}^{50}}{C_{100}^{50}} + \frac{C_5^1 \cdot C_{95}^{49}}{C_{100}^{50}}.$$

**Bài 21:** Chọn hú họa ra một quân cờ tướng từ một bộ cờ gồm 32 quân. Gọi  $A$  là biến cố rút được quân tướng, còn  $B$  là biến cố rút ra quân cờ đen. Hỏi  $A$  và  $B$  có độc lập không?

**Giải:** Chú ý bộ cờ tướng có 32 quân gồm 16 quân đen, 16 quân trắng, trong đó có 2 tướng (đen và trắng). Dễ thấy:

$$P(A) = \frac{2}{32} = \frac{1}{16}; \quad P(B) = \frac{16}{32} = \frac{1}{2}.$$



Mặt khác biến cố rút được một con tướng đen sẽ là AB và  $P(AB) = \frac{1}{32} = P(A)P(B)$  nên A và B là các biến cố độc lập.

**Bài 22:** Xác suất để một thiết bị bị trục trặc trong một ngày làm việc là  $\alpha=0,01$ . Tìm xác suất để trong vòng 5 ngày máy làm việc tốt.

**Giải:** Ta coi thiết bị làm việc trong ngày này độc lập so với ngày khác, như vậy xác suất để nó làm được tốt trong 5 ngày sẽ bằng tích xác suất làm việc tốt trong từng ngày, từ đó xác suất cần tìm là:

$$(1 - \alpha)^5 = (1 - 0,01)^5 \approx 1 - 5\alpha = 0,95$$

(do  $\alpha$  rất bé).

## Chương 2

# CÔNG THỨC XÁC SUẤT TOÀN PHẦN VÀ CÔNG THỨC BÂYẾT. DÃY PHÉP THỬ ĐỘC LẬP

### 2.1. CÔNG THỨC XÁC SUẤT TOÀN PHẦN

Một hệ quả quan trọng của định lý cộng xác suất và định lý nhân xác suất là công thức xác suất toàn phần (hay công thức xác suất đầy đủ).

#### 2.1.1. Nhóm đầy đủ các biến cố xung khắc

*a) Định nghĩa:*

Các biến cố  $A_1, A_2, \dots, A_n$  được gọi là hợp thành một nhóm đầy đủ các biến cố xung khắc nếu chúng xung khắc từng đôi một và nhất thiết, một trong chúng xảy ra khi thực hiện phép thử, nghĩa là :

$$A_i A_j = V \text{ (biến cố không thể có)}$$

$$\text{với } i \neq j \text{ và } i, j = 1, 2, \dots, n$$

$$A_1 + A_2 + \dots + A_n = U \text{ (biến cố chắc chắn)}$$

*b) Ví dụ:*

Gọi  $A_i$  ( $i=1, 2, 3, 4, 5, 6$ ) là biến cố xuất hiện mặt  $i$  chấm trong phép thử gieo một con xúc xắc. Khi đó, các biến cố  $A_i$  hợp thành một nhóm đầy đủ các biến cố xung khắc. Thực vậy, khi gieo một con xúc xắc ta thấy không cùng xảy ra mặt 1 chấm và mặt 2 chấm, mặt 1 chấm và mặt 3 chấm, ... , nghĩa là chúng xung khắc từng đôi một và ngoài ra ta cũng thấy nhất thiết mặt có 1 chấm hoặc mặt có 2 chấm



hoặc mặt có 3 chấm, ... xuất hiện, nghĩa là nhất thiết một trong chúng xảy ra.

### 2.1.2. Lập công thức

Giả sử  $A_1, A_2, \dots, A_n$  là một nhóm đầy đủ các biến cố xung khắc và  $B$  là một biến cố bất kỳ nào đó khác cơ chế xảy ra và luôn xảy ra cùng với  $A_i$  trong một phép thử ( $i=1, 2, \dots, n$ ). Cho biết xác suất  $P(A_i)$  và  $P_{A_i}(B)$ . Hãy tính xác suất của biến cố  $B$  theo  $P(A_i)$  và  $P_{A_i}(B)$ . Ta tìm công thức tính xác suất  $P(B)$  như sau:

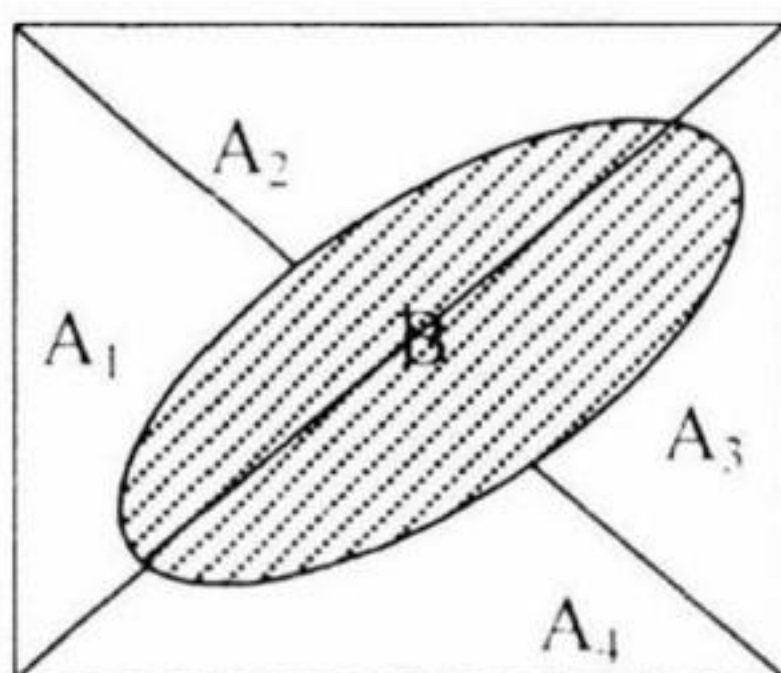
Theo giả thiết, rõ ràng ta có:

$$B = A_1 B + A_2 B + \dots + A_n B \quad (1)$$

(bạn đọc có thể hình dung trực quan bởi hình 8).

Vì các biến cố  $A_i$  xung khắc nên các biến cố  $A_i B$  ( $i=1, 2, \dots, n$ ) cũng xung khắc, do đó theo qui tắc cộng xác suất ta có :

$$P(B) = \sum_{i=1}^n P(A_i B) \quad (2)$$



**Hình 8**

Hơn nữa theo định lý nhân xác suất:

$$P(A_i B) = P(A_i) P_{A_i}(B) \quad (3)$$

Thành thử:

$$P(B) = \sum_{i=1}^n P(A_i) P_{A_i}(B) \quad (4)$$

Công thức (4) cho phép tính xác suất của B theo các xác suất  $P(A_i)$  và các xác suất có điều kiện  $P_{A_i}(B)$  ( $i=1, 2, \dots, n$ ). Đó cũng là công thức xác suất toàn phần.

### 2.1.3. Các ví dụ áp dụng công thức xác suất toàn phần

*Ví dụ 1:* Một nhà máy sản xuất bóng đèn gồm 3 phân xưởng: phân xưởng I sản xuất 50%, phân xưởng II sản xuất 20%, phân xưởng III sản xuất 30% tổng số bóng đèn của toàn nhà máy. Tỷ lệ phế phẩm tương ứng của các phân xưởng là 2%; 3% và 4%. Hãy tính tỷ lệ phế phẩm chung của toàn nhà máy.

*Giải:* Tỷ lệ phế phẩm cần tìm chính là xác suất để một bóng đèn chọn hú họa từ kho sản phẩm của nhà máy là phế phẩm.

Gọi  $A_i$  ( $i=1, 2, 3$ ) là biến cố đèn chọn ra thuộc phân xưởng  $i$  ( $i=I, II, III$ ).

Gọi B là biến cố bóng đèn chọn ra là phế phẩm.

Rõ ràng các biến cố  $A_i$  ( $i=1, 2, 3$ ) hợp thành nhóm đầy đủ các biến cố. Theo giả thiết ta có:

$$P(A_1) = 0,5; P(A_2) = 0,2; P(A_3) = 0,3;$$

$$P_{A_1}(B) = 0,02; P_{A_2}(B) = 0,03; P_{A_3}(B) = 0,04.$$

Do đó:

$$P(B) = P(A_1).P_{A_1}(B) + P(A_2).P_{A_2}(B) + P(A_3).P_{A_3}(B)$$

$$= 0,5.0,02 + 0,2.0,03 + 0,3.0,04 = 0,028 = 2,8\%$$

*Ví dụ 2:* Cho biết xác suất để trong khoảng thời gian  $t$  có  $k$  lần gọi đến một tổng đài điện thoại là  $P_t(k)$ . Giả sử rằng biến cố  $k$  lần gọi trong khoảng thời gian này và biến cố  $l$  lần gọi trong khoảng thời gian khác không giao với khoảng trên là các biến cố độc lập với mọi  $k, l$  bất kỳ.

Hãy tính xác suất  $P_{2t}(n)$  của biến cố: trong thời gian  $2t$  có  $n$  lần gọi.

*Giải:* Gọi B là biến cố có  $n$  lần gọi trong khoảng thời gian  $2t$ . Gọi  $A_k$  là biến cố có  $k$  lần gọi trong nửa khoảng đầu của khoảng thời gian



2t này :  $P(A_k) = P_1(k)$  ( $k=0, 1, 2, \dots, n$ ). Rõ ràng để biến cố xuất hiện với điều kiện trong nửa khoảng thời gian đầu  $A_k$  đã xảy ra là trong nửa khoảng thời gian còn lại phải xuất hiện  $n-k$  lần gọi. Do đó

$$P_{A_k}(B) = P_1(n-k)$$

Áp dụng công thức xác suất toàn phần ta có:

$$P(B) = P_{2t}(n) = \sum_{k=0}^n P_1(k)P_1(n-k)$$

## 2.2. CÔNG THỨC BÂYÉT

### 2.2.1. Lập công thức

Với cùng các giả thiết như khi thiết lập công thức xác suất toàn phần trong mục 2.1 và ta nêu thêm một điều kiện mới: phép thử được thực hiện và kết quả là biến cố xảy ra. Vì các  $A_i$  ( $i = 1, 2, \dots, n$ ) hợp thành nhóm đầy đủ nên cùng với  $B$  xuất hiện thì biến cố  $A_i$  nào đó xảy ra (và chỉ một). Thử hỏi khi  $B$  xuất hiện thì biến cố nào trong số các biến cố  $A_i$  có nhiều khả năng xuất hiện hơn cả? Giải đáp câu hỏi này có nghĩa là cần tính các xác suất có điều kiện  $P_B(A_i)$  ( $i=1, 2, \dots, n$ ).

Theo qui tắc nhân xác suất ta có:

$$P(B) \cdot P_B(A_i) = P(A_i) \cdot P_{A_i}(B) \quad (5)$$

Suy ra:

$$P_B(A_i) = \frac{P(A_i) \cdot P_{A_i}(B)}{P(B)} \quad (6)$$

Thay  $P(B)$  bởi công thức (4) trong 2.1 ta được:

$$P_B(A_i) = \frac{P(A_i) \cdot P_{A_i}(B)}{\sum_{i=1}^n P(A_i) \cdot P_{A_i}(B)} \quad (7)$$

Công thức (7) gọi là công thức Bâyét.

Các xác suất  $P_B(A_i)$  thường được gọi là *các xác suất hậu nghiệm* để phân biệt với *các xác suất tiên nghiệm*  $P(A_i)$ .

### 2.2.2. Các ví dụ áp dụng công thức Bayes

*Ví dụ 1:* Một thiết bị gồm 3 loại linh kiện: loại I chiếm 35%, loại II chiếm 25%, loại III chiếm 40% tổng số linh kiện của toàn thiết bị. Xác suất hư hỏng sau khoảng thời gian công tác nào đó của các loại tương ứng là: 15%, 25% và 5%.

Máy đang hoạt động bỗng bị hỏng, hãy tính xác suất để từng loại linh kiện bị hỏng (giả thiết các loại linh kiện không cùng hỏng đồng thời).

*Giải:* Gọi B là biến cố thiết bị ngừng hoạt động. Ta phải tính các xác suất  $P_B(A_i)$  tức là xác suất để có sự cố ở linh kiện loại i (i=I, II, III).

Theo công thức (7) ta có:

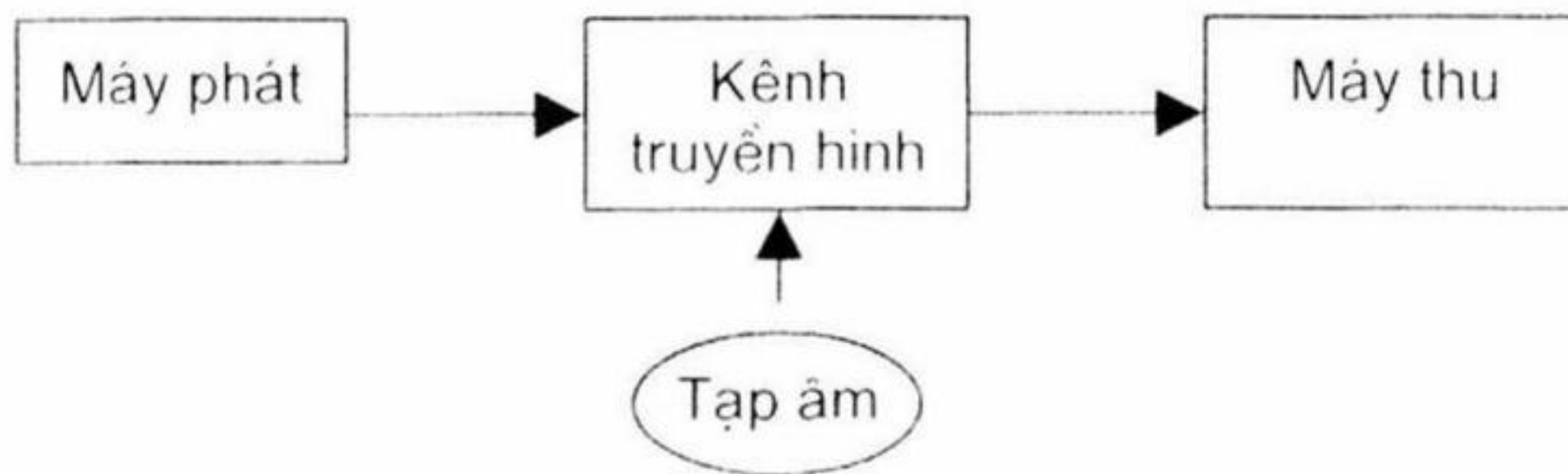
$$P_B(A_1) = \frac{0,35 \cdot 0,15}{0,35 \cdot 0,15 + 0,25 \cdot 0,25 + 0,40 \cdot 0,05} = \frac{525}{1550} = \frac{21}{62};$$

$$P_B(A_2) = \frac{25}{62}; \quad P_B(A_3) = \frac{16}{62}.$$

Việc tính các xác suất này có ý nghĩa thực tiễn trong việc sử dụng và thiết kế: tuy số linh kiện loại II chỉ chiếm số ít trong tổng số linh kiện của máy nhưng vì chúng hay hỏng nên khi thấy máy hỏng thì trước tiên nên kiểm tra các linh kiện loại II, do vậy khi lắp máy nên bố trí các linh kiện loại này ở vị trí dễ tháo lắp nhất.

*Ví dụ 2:* Có một hệ thống truyền tin như ở hình 9. Tại máy phát có thể xảy ra một trong hai biến cố: phát tín hiệu (biến cố B) và không phát tín hiệu (biến cố  $\bar{B}$ ). Tại máy thu cũng có thể xảy ra một trong hai biến cố: nhận được tín hiệu (biến cố A) và không nhận được tín hiệu (biến cố  $\bar{A}$ ). Vì ảnh hưởng can nhiễu của tạp âm lên kênh truyền tin nên có thể xảy ra hiện tượng ở máy phát có tín hiệu phát đi mà máy thu không nhận được hoặc ngược lại máy phát không phát mà máy thu vẫn nhận được tín hiệu (tín hiệu giả do tạp âm gây ra).





**Hình 9**

Để xác định độ tin cậy của hệ thống truyền tin, cần tính các xác suất  $P_A(B)$  và  $P_A(\bar{B})$  (xác suất để thật sự có tín hiệu phát đi khi ở máy thu nhận được tín hiệu và xác suất để thật sự không có tín hiệu phát đi khi máy thu không nhận được tín hiệu).

Giả sử đã biết các xác suất  $P(B)$ ,  $P_B(A)$ ,  $P_B(\bar{A})$ . Khi đó theo công thức Bâyét ta có:

$$P_A(B) = \frac{P(B) \cdot P_B(A)}{P(B) \cdot P_B(A) + P(\bar{B}) \cdot P_B(\bar{A})}$$

$$P_A(\bar{B}) = \frac{P(\bar{B}) \cdot P_B(\bar{A})}{P(\bar{B}) \cdot P_B(\bar{A}) + P(B) \cdot P_B(A)}$$

Áp dụng bằng số: với  $P(B) = \frac{5}{8}$ ,  $P(\bar{B}) = \frac{3}{8}$ ,  $P_B(A) = \frac{3}{5}$ ,  $P_B(\bar{A}) = \frac{1}{3}$ .

Ta tính ra:

$$P_A(B) = \frac{3}{4}; \quad P_A(\bar{B}) = \frac{1}{2}.$$

## 2.3. CÔNG THỨC BÉC-NU-LI CHO DÃY PHÉP THỬ ĐỘC LẬP

### 2.3.1. Lập công thức

Trong thực tế ta thường gặp những phép thử phức hợp tạo thành từ những phép thử như nhau lặp đi lặp lại, và ta phải tính xác suất của các kết cục khác nhau của phép thử này.

Ta nói rằng các phép thử là độc lập nếu xác suất của một kết cục bất kỳ của mỗi phép thử không phụ thuộc vào kết cục của các phép thử khác.

Bài toán đặt ra như sau: Giả sử tiến hành  $n$  phép thử lặp đi lặp lại trong cùng những điều kiện như nhau, mỗi phép thử có hai kết cục  $A$  và  $\bar{A}$ . Xác suất xuất hiện sự kiện  $A$  ở mỗi phép thử đều bằng  $p$  (không phụ thuộc chỉ số phép thử). Hãy tính xác suất sao cho khi thực hiện  $n$  phép thử như vậy sự kiện  $A$  xuất hiện  $k$  lần ( $k=0, 1, 2, \dots, n$ ).

Nhà toán học Thụy Sĩ I-a-côp Béc-nu-li đã giải được bài toán trên từ thế kỷ 17.

Trước hết chú ý rằng vì  $\bar{A}$  là sự kiện đối lập của  $A$  nên xác suất xuất hiện sự kiện  $\bar{A}$  trong mỗi phép thử là  $P(\bar{A}) = 1 - p = q$  (đặt  $q = 1 - p$ ).

Ký hiệu  $P_n(k)$  là xác suất để trong  $n$  phép thử độc lập, sự kiện  $A$  xuất hiện  $k$  lần (xác suất của sự kiện  $B$ ). Sự kiện  $B$  có thể thực hiện theo nhiều cách khác nhau: hoặc trong  $k$  phép thử đầu  $A$  xuất hiện và  $n - k$  phép thử còn lại  $\bar{A}$  xuất hiện (ký hiệu  $A_1 A_2 \dots A_k \bar{A}_{k+1} \dots \bar{A}_n$ ), ... , hoặc trong  $n - k$  phép thử đầu  $\bar{A}$  xuất hiện và  $k$  phép thử còn lại  $A$  xuất hiện. Mỗi cách này ứng với việc chọn ra  $k$  phép thử từ  $n$  phép thử đã cho, vì vậy có tất cả  $C_n^k$  (số tổ hợp) cách. Theo giả thiết các phép thử là độc lập và xác suất xuất hiện sự kiện  $A$  không đổi nên ta có:

$$P(A_1 A_2 \dots A_k \bar{A}_{k+1} \dots \bar{A}_n) = \underbrace{p \cdot p \dots p}_{k \text{ lần}} \cdot \underbrace{q \dots q}_{n-k \text{ lần}} = p^k q^{n-k}$$

Đó cũng là xác suất của mỗi một trong  $C_n^k$  cách trên. Các cách đó rõ ràng xung khắc với nhau nên theo định lý cộng xác suất ta có:

$$P(B) = P_n(k) = C_n^k p^k q^{n-k}$$

Đây là công thức Béc-nu-li.

### 2.3.2. Các ví dụ áp dụng công thức Béc-nu-li

**Ví dụ 1:** Xác suất tiêu thụ điện năng trong mỗi ngày không vượt mức quy định tại một xí nghiệp là  $p=0,75$ . Tính xác suất sao cho trong 6 ngày liên tiếp có 4 ngày lượng điện năng tiêu thụ không vượt mức quy định.

**Giải:** Ta có  $n = 6$ ,  $k = 4$ ,  $p = P(A) = 0,75$ ,  $q = 0,25$ .



Xác suất phải tìm là:

$$P_6(4) = C_6^4 (0,75)^4 (0,25)^2 = 0,30.$$

**Ví dụ 2:** Một gia đình có 10 người con. Coi xác suất sinh con trai bằng xác suất sinh con gái, hãy tính xác suất để trong gia đình đó có:

- 5 con trai.
- Số con trai không dưới 3 và không quá 8.

**Giải:** Gọi  $p$  là xác suất sinh con trai, thì  $p = 1/2$  và xác suất sinh con gái  $q = 1 - p = 1/2$ .

a) Xác suất để gia đình có 5 con trai được tính theo công thức Béc-nu-li (1):

$$P_{10}(5) = C_{10}^5 p^5 q^{10-5} = C_{10}^5 \left(\frac{1}{2}\right)^{10} = \frac{63}{256}.$$

b) Gọi  $A$  là biến cố gia đình đó có số con trai dưới 3 (không, một, hai con trai) và quá 8 (9, 10 con trai). Khi đó:

$$P(A) = [C_{10}^0 p^0 q^{10} + C_{10}^1 p^1 q^9 + C_{10}^2 p^2 q^8] + [C_{10}^9 p^9 q^1 + C_{10}^{10} p^{10} q^0] = \frac{67}{1024}$$

Khi đó là  $\bar{A}$  biến cố gia đình đó có số con trai không dưới 3 và không quá 8, ta có:

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{67}{1024} = \frac{957}{1024}.$$

## 2.4. CÁC ĐỊNH LÝ GIỚI HẠN MOA-VƠ-RƠ LAPLAT

### 2.4.1. Định lý 1

Nếu trong mỗi phép thử độc lập sự kiện  $A$  xuất hiện với xác suất  $p$  ( $0 < p < 1$ ) thì khi  $n \rightarrow \infty$  ta có:

$$\lim_{n \rightarrow \infty} \left[ P_n(k) - \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \right] = 0 \quad (8)$$

Điều đó có nghĩa là với  $n$  khá lớn thì:

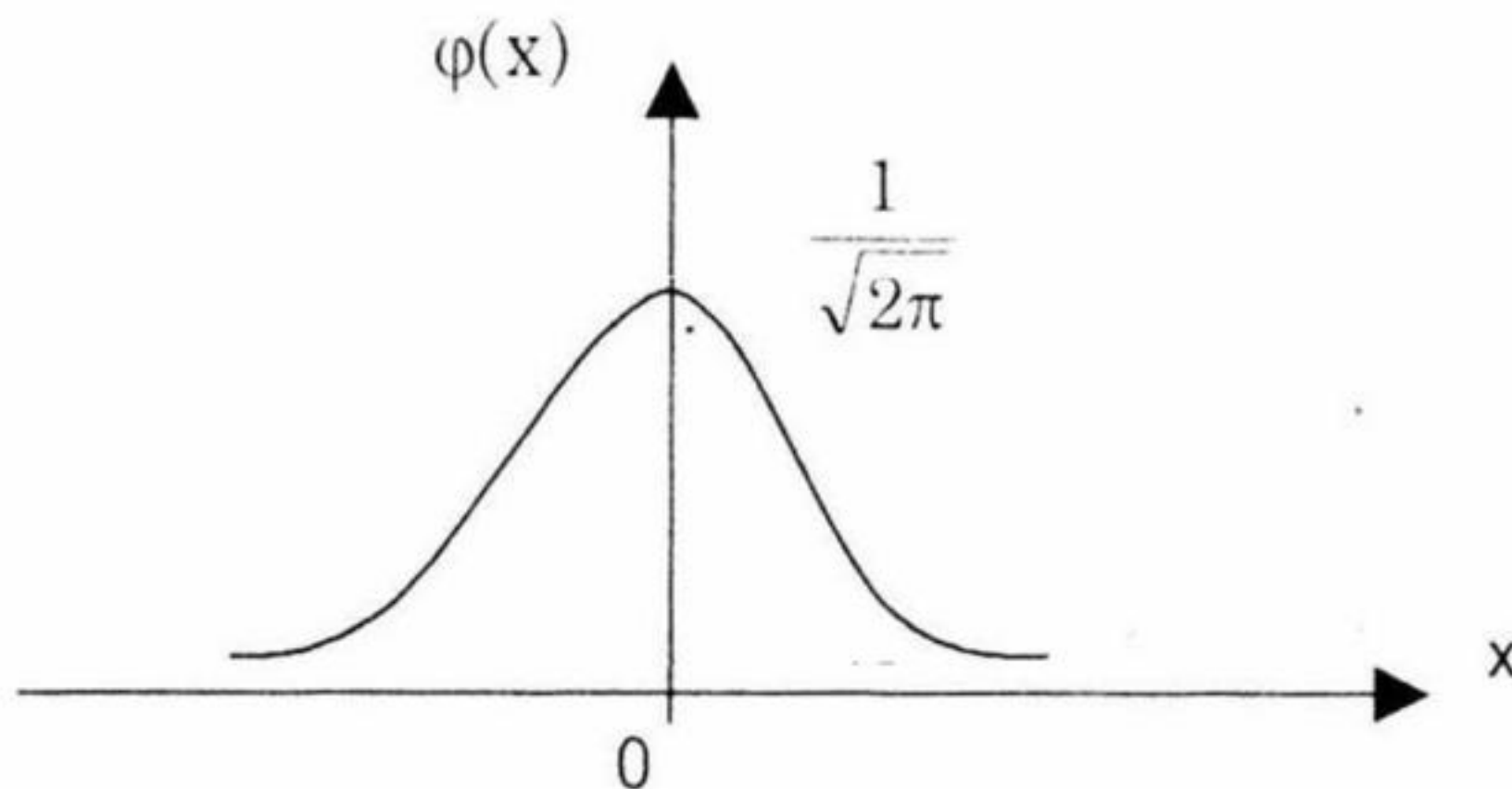
$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x_0) \quad (9)$$

trong đó  $\varphi(x_0)$  là giá trị của hàm số Gauxơ  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  tại trị số của đối số  $x = x_0 = \frac{k - np}{\sqrt{npq}}$ .

Hàm số  $\varphi(x)$  đã được lập bảng (xem bảng I ở cuối sách) ứng với các trị của đối số từ  $x=0$  đến  $x=3,9$ . Khi  $x>3,9$  hàm này có giá trị rất gần 0 và giảm rất chậm do đó có thể lấy  $\varphi(3,9)$  làm giá trị của  $\varphi(x)$  tại mọi  $x>3,9$ .

Để tính trị số của  $\varphi(x)$  tại  $x<0$  ta chú ý rằng  $\varphi(x)$  là hàm số chẵn  $\varphi(x) = \varphi(-x)$ .

Đồ thị của  $\varphi(x)$  có dạng hình chuông và thường được gọi là đường cong Gauxơ (hình 10).



**Hình 10**

**Ví dụ 1:** Tính xác suất để trong dãy gồm 400 phép thử độc lập, sự kiện A xuất hiện 80 lần, nếu mỗi phép thử sự kiện A xuất hiện với xác suất  $p = 0,2$ .

**Giải:** Ta có  $n=400$ ;  $k=80$ ;  $p=0,2$ ;  $q=0,8$ .



$$P_{100}(80) \approx \frac{1}{\sqrt{400 \cdot 0,2 \cdot 0,8}} \varphi(x_0)$$

trong đó:

$$x_0 = \frac{k - np}{\sqrt{npq}} = \frac{80 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = 0$$

Tra bảng ta có  $\varphi(x_0) = \varphi(0) \approx 0,3989$ ; do đó:

$$P_{100}(80) \approx \frac{1}{8} \cdot 0,3989 \approx 0,05,$$

## 2.4.2. Định lý 2

Với các điều kiện như ở định lý 1 ta có:

$$\lim_{n \rightarrow \infty} \left[ P_n(k_1; k_2) - \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt \right] = 0 \quad (10)$$

trong đó: 
$$x_i = \frac{k_i - np}{\sqrt{npq}} \quad (i=1, 2)$$

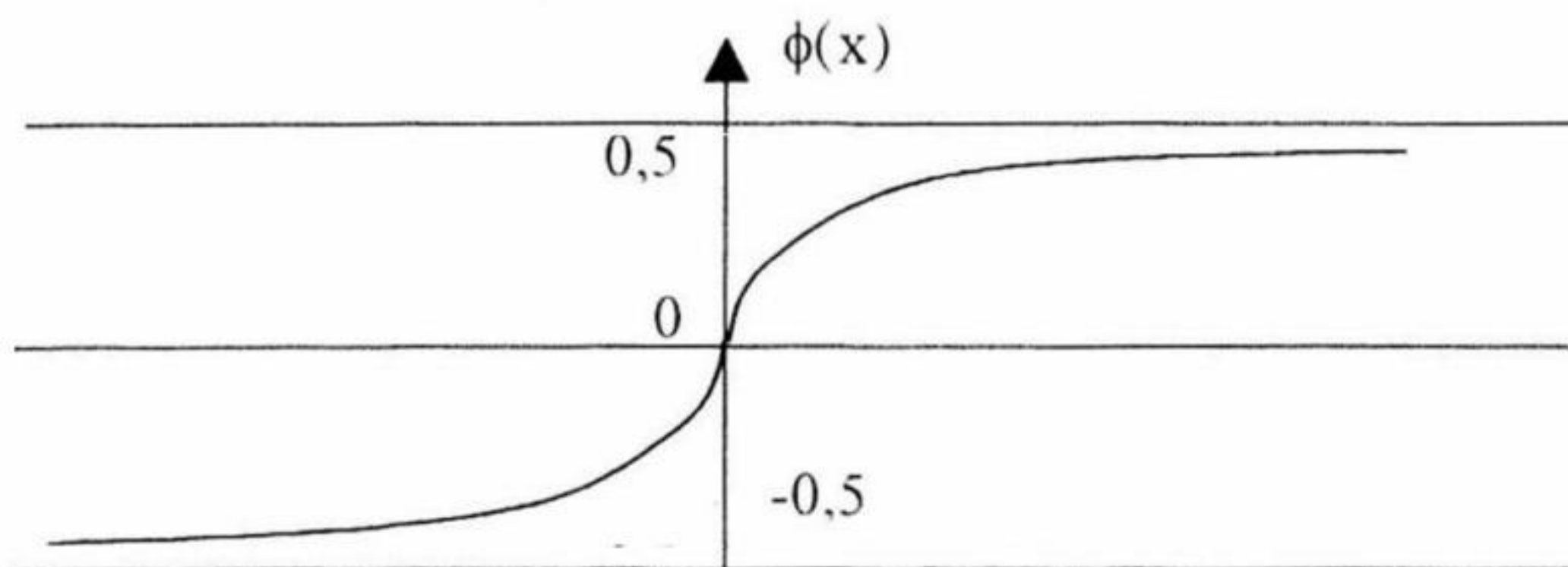
Điều đó có nghĩa là với  $n$  khá lớn ta có công thức xấp xỉ:

$$P_n(k_1; k_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt = \Phi(x_2) - \Phi(x_1) \quad (11)$$

trong đó  $\Phi(x)$  là hàm số Laplát:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad (12)$$

Để tính gần đúng xác suất  $P_n(k_1; k_2)$  theo công thức (11) người ta đã lập bảng giá trị của hàm  $\phi(x)$  ứng với phạm vi biến thiên của  $x$  từ 0 đến 5 cách nhau một phần mười đơn vị (xem bảng II cuối sách). Để tính giá trị của  $\phi(x)$  khi  $x < 0$  ta chú ý tính chất lẻ của  $\phi$ :  $\phi(-x) = -\phi(x)$ . Còn với  $x > 5$  hàm  $\phi(x)$  tăng rất chậm nên có thể lấy  $\phi(5)$  làm giá trị của hàm tại mỗi  $x > 5$ . Đồ thị của  $\phi(x)$  có dạng như hình 11.



**Hình 11**

Các công thức xấp xỉ (10) và (11) sẽ cho kết quả càng chính xác khi  $n$  càng lớn và  $p$  càng gần  $1/2$ . Khi  $p$  quá bé hoặc quá lớn thì sai số mắc phải sẽ lớn.

Các định lý trên được Moavơ xây dựng năm 1730 cho trường hợp riêng  $p=q=1/2$ , và đến năm 1883 được Laplat tổng quát hoá với  $p$  có trị số bất kỳ ( $0 < p < 1$ ). Đó là trường hợp riêng của một định lý tổng quát được gọi là định lý giới hạn trung tâm trong lý thuyết xác suất.

*Ví dụ 2:* Xác suất bắn trúng đích của một xạ thủ là  $p=0,75$ . Tính xác suất để với 100 phát có 81 phát trúng đích trở lên.

*Giải:* Theo (11) ta có:

$$P_{100}(81;100) \approx \phi(x_2) - \phi(x_1)$$

trong đó:

$$x_1 = \frac{81 - 100 \cdot 0,75}{\sqrt{100 \cdot 0,75 \cdot 0,25}} \approx 1,38$$

$$x_2 = \frac{100 - 100 \cdot 0,75}{\sqrt{100 \cdot 0,75 \cdot 0,25}} \approx 5,77$$

Tra bảng ta có:  $\phi(1,38) = 0,4162$

$$\phi(5,77) = \phi(5) = 0,5000$$

Vậy  $P_{100}(81; 100) = \phi(5,77) - \phi(1,38)$

$$= 0,5000 - 0,4162 = 0,0838$$



## 2.5. ĐỊNH LÝ GIỚI HẠN POÁTXÔNG

Như đã nói trên công thức Bécnuili chỉ áp dụng trên thực tế khi số phép thử tương đối bé. Trường hợp  $n$  lớn và  $p$  không quá gần 0 hay 1 ta dùng các công thức xấp xỉ (9) và (11) trong 2.4 để tính  $P_n(k)$  và  $P_n(k_1; k_2)$ . Trong trường hợp  $p$  quá gần 0 hoặc 1 thì việc dùng các công thức đó sẽ dẫn tới sai số đáng kể. Trong trường hợp  $n$  rất lớn và  $p$  rất bé ta tính  $P_n(k)$  nhờ định lý: Định lý Poátxông

### 2.5.1. Định lý Poátxông

Giả sử tiến hành  $n$  phép thử độc lập, mỗi phép thử sự kiện  $A$  xuất hiện với xác suất  $P(A) = p$ .

Nếu khi  $n \rightarrow \infty$  mà  $p \rightarrow 0$  sao cho  $np = \lambda = \text{const}$  thì ta có:

$$\lim_{n \rightarrow \infty} P_n(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (13)$$

*Chứng minh:* Theo công thức Bécnuili :

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

Theo giả thiết  $p = \frac{\lambda}{n}$ , do đó:

$$\begin{aligned} P_n(k) &= \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left[1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{(k-1)}{n}\right)\right] \times \left(1 - \frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

Chuyển qua giới hạn khi  $n \rightarrow \infty$  ta được ( $k$  cố định):

$$\lim_{n \rightarrow \infty} P_n(k) = \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Từ công thức (13) ta có công thức xấp xỉ sau đây ( $n$  lớn,  $p$  bé và  $np=\lambda$ ).

$$P_n(k) \approx e^{-\lambda} \frac{\lambda^k}{k!} \quad (14)$$

Vì các sự kiện có xác suất bé sẽ rất ít khả năng xảy ra nên định lý Poátxông thường được gọi là định luật về các sự kiện hiếm hoi.

*Chú thích:* Trong trường hợp  $p$  rất gần 1 thì  $P(A)=1-p$  sẽ rất gần 0 do đó để tính  $P_n(k)$  ta có thể chuyển sang tính xác suất sao cho trong  $n$  phép thử  $\bar{A}$  xảy ra  $n-k$  lần.

Để tiện cho áp dụng, trong các sách về lý thuyết xác suất người ta thường tính sẵn giá trị của hàm  $e^{-\lambda} \frac{\lambda^k}{k!}$  ứng với các cặp trị số của  $\lambda$  và  $k$ , cũng như bảng giá trị  $\sum_{k=0}^m e^{-\lambda} \frac{\lambda^k}{k!}$  để tính các xác suất  $P_n(k_1; k_2)$  (xem bảng 3 và 4 phần phụ lục).

### 2.5.2. Ví dụ

Một nữ công nhân đứng máy xe sợi gồm 800 ống sợi, xác suất đứt sợi ở mỗi ống trong một giờ là 0,005. Tính xác suất của các sự kiện:

- Trong vòng 1 giờ có 3 ống bị đứt sợi;
- Trong vòng 1 giờ có không quá 10 ống bị đứt sợi.

Giải: Ở đây  $n = 800$ ;  $p = 0,005$ ;  $\lambda = np = 4$ .

Theo công thức (14), ta có:

$$a) P_{800}(3) \approx \frac{4^3 e^{-4}}{3!} = 0,1954 \quad (\text{tra bảng 3})$$

$$b) P_{800}(0;10) = \sum_{k=0}^{10} P_{800}(k) = 0,99716 \quad (\text{tra bảng 4}).$$

## 2.6. SỐ LẦN XUẤT HIỆN CHẮC NHẤT

Trị số của xác suất  $P_n(k)$  nói chung phụ thuộc vào  $k$  (với  $n$  cố định). Ta muốn tìm trị số  $k = k_0$  sao cho  $P_n(k_0)$  đạt giá trị lớn nhất. Số  $k_0$  được gọi là số lần xuất hiện chắc nhất của sự kiện  $A$  trong dãy  $n$  phép thử đã cho.



Dĩ nhiên có thể thu được lời giải bằng cách tính tất cả các trị số của  $P_n(k)$  ứng lần lượt với  $k = 0, 1, 2, \dots, n$ , nhưng cách làm “thủ công” đó tốn rất nhiều thời gian và không tổng quát. Vì vậy điều hợp lý là ta sẽ xem  $P_n(k)$  như là hàm của đối số tự nhiên  $k$ , xét đáng điệu biến thiên của  $P_n(k)$  theo  $k$  rồi từ đó tìm ra  $k_0$ .

Nhằm mục đích này ta lập tỷ số:

$$\begin{aligned} \frac{P_n(k+1)}{P_n(k)} &= \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \\ &= \frac{n!}{(k+1)!(n-k-1)!} \cdot \frac{k!(n-k)!}{n!} \cdot \frac{p}{q} = \frac{n-k}{k+1} \cdot \frac{p}{q} \end{aligned}$$

Từ đó suy ra rằng  $P_n(k+1) > P_n(k)$  khi  $(n-k)p > (k+1)q$  tức là khi  $k < np - q$ ;  $P_n(k+1) = P_n(k)$ , khi  $k = np - q$  và cuối cùng:  $P_n(k+1) < P_n(k)$ , khi  $k > np - q$ .

Ta thấy khi  $k$  tăng từ 0 đến  $n$ , hàm  $P_n(k)$  thoạt tiên tăng, sau đó đạt cực đại rồi giảm dần.

Nếu  $np - q$  là số nguyên thì  $P_n(k)$  đạt cực đại tại 2 giá trị của  $k$  là:  $k_0 = np - q$  và  $k_0 + 1 = np - q + 1 = np + p = (n+1)p$ .

Còn nếu  $np - q$  không nguyên thì  $P_n(k)$  đạt cực đại tại  $k = k_0$  với  $k_0$  là số nguyên bé nhất lớn hơn  $np - q$  nghĩa là phần nguyên của số  $np - q + 1 = np + p = p(n+1)$ .

Đặc biệt nếu  $np - q < 0$  thì theo trên  $P_n(k)$  sẽ luôn luôn giảm:  $P_n(0) > P_n(1) > \dots > P_n(n)$ , còn nếu  $np - q = 0$  thì:

$$P_n(0) = P_n(1) > P_n(2) > \dots > P_n(n).$$

Số  $k_0$  được gọi là số lần xuất hiện chắc nhất hay trị số chắc nhất của  $k$ .

Ví dụ: Giả sử  $n = 50$ ,  $p = 1/3$ .

Ta có:

$$np - q = \frac{50}{3} - \frac{2}{3} = \frac{48}{3} = 16$$

Vì  $np - q = 16$  là số nguyên không âm, vậy có hai giá trị chắc nhất là 16 và 17, tại đây xác suất  $P_{50}(k)$  đạt giá trị cực đại:

$$P_{50}(16) = P_{50}(17) = 0,1178$$

## 2.7. ĐỊNH LÝ TỔNG QUÁT VỀ CÁC PHÉP THỬ LẬP

Công thức Bécnuili cho ta tính xác suất  $P_n(k)$  với giả thiết xác suất của sự kiện trong mỗi phép thử đều bằng  $p$ . Trong trường hợp đó các phép thử được lập đi lập lại dưới cùng bộ điều kiện như nhau.

Bây giờ ta giả sử xác suất của sự kiện  $A$  phụ thuộc vào chỉ số phép thử, tức là các phép thử được tiến hành dưới các điều kiện khác nhau.

Gọi  $p_i$  là xác suất xuất hiện  $A$  ở phép thử  $i$  ( $i=1, 2, \dots, n$ ). Hãy tính xác suất để trong  $n$  phép thử độc lập sự kiện  $A$  xảy ra  $k$  lần.

Gọi  $B$  là sự kiện phức hợp nói trên. Ta biểu diễn  $B$  dưới dạng:

$$B = A_1 A_2 \cdots A_k \bar{A}_{k+1} \cdots + \bar{A}_1 \bar{A}_2 \cdots \bar{A}_{k-k} A_{k-k+1} \cdots A_n$$

Số các sự kiện thành phần trong tổng (sự kiện) này là  $C_n^k$ , tuy nhiên trong trường hợp này chúng không cùng xác suất như lược đồ Bécnuili ở 2.1.2.

Áp dụng định lý cộng và định lý nhân xác suất ta được:

$$P(B) = p_1 p_2 \cdots p_k q_{k+1} \cdots q_n + \cdots + q_1 q_2 \cdots q_{n-k} p_{n-k+1} \cdots p_n \quad (15)$$

Nghĩa là xác suất phải tìm bằng tổng tất cả các tích có thể tạo nên từ hai chữ  $p$  và  $q$  trong đó chữ  $p$  với các chỉ số khác nhau có mặt  $k$  lần, còn chữ  $q$  với các chỉ số khác nhau có mặt  $n-k$  lần.

Có thể thu được nhanh chóng xác suất  $P(B)$  bằng phương pháp lập hàm sinh, nghĩa là lấy hàm số

$$\varphi_n(z) = \prod_{i=1}^n (q_i + p_i z) \quad (16)$$

Ta hãy tìm hệ số của số hạng bậc nhất đối với  $z$  trong khai triển của (16). Rõ ràng một số hạng chứa  $z^k$  sẽ có hệ số là tích của  $k$  chữ  $p$  với các chỉ số khác nhau và  $n-k$  chữ  $q$  với các chỉ số khác nhau, thành



thứ sau khi gộp tất cả các số hạng ấy lại ta được hệ số của  $z^k$  trong dạng (15). Ta có định lý sau:

**Định lý:** Xác suất để trong  $n$  phép thử độc lập, sự kiện  $A$  xuất hiện  $k$  lần bằng hệ số của số hạng chứa  $z^k$  trong khai triển của hàm sinh:

$$\varphi_n(z) = \prod_{i=1}^n (q_i + p_i z)$$

(trong đó  $p_i$  là xác suất của sự kiện  $A$  ở phép thử  $i$ ; còn  $q_i = 1 - p_i$ ).

Đặc biệt khi  $p_1 = p_2 = \dots = p_n = p$ , thì từ (15) ta được :

$$P(B) = P_n(k) = C_n^k p^k q^{n-k}$$

đó chính là công thức Bécnu-li quen biết.

*Ví dụ:* Bắn 4 phát súng liên tiếp vào một mục tiêu di động, xác suất tương ứng của mỗi phát là:  $p_1=0,1$ ;  $p_2=0,2$ ;  $p_3=0,3$ ;  $p_4=0,4$  (có thể hình dung mục tiêu tiến dần đến xạ thủ).

Tính xác suất của các sự kiện:

- a) Không có phát nào trúng;
- b) Có một phát trúng;
- c) Có hai phát trúng;
- d) Có ba phát trúng;
- e) Có bốn phát trúng.

*Giải:* Lập hàm sinh:

$$\varphi_4(z) = \prod_{i=1}^4 (q_i + p_i z) = (0,9 + 0,1z)(0,8 + 0,2z)(0,7 + 0,3z)(0,6 + 0,4z).$$

Khai triển tích này ta có:

$$\varphi_4(z) = 0,302 + 0,440z + 0,215z^2 + 0,040z^3 + 0,002z^4$$

Do đó:

- a)  $P_4(0)=0,302$ ;
- b)  $P_4(1)=0,440$ ;
- c)  $P_4(2)=0,215$ ;
- d)  $P_4(3)=0,040$ ;
- e)  $P_4(4)=0,002$ .

## BÀI TẬP CHƯƠNG 2

**Bài 1:** Để sản xuất một loại sản phẩm, có thể dùng một trong hai máy. Tỷ lệ phế phẩm đối với máy thứ nhất là 0,03 và đối với máy thứ hai là 0,02. Trong một kho gồm  $\frac{2}{3}$  sản phẩm của máy thứ nhất và  $\frac{1}{3}$  sản phẩm của máy thứ hai, người ta lấy ra một sản phẩm (lấy hù họa). Tính xác suất để sản phẩm đó:

- a) Là phế phẩm?
- b) Không là phế phẩm?

**Giải:**

a) Gọi  $A_1$  và  $A_2$  là biến cố lấy ra một sản phẩm của máy thứ nhất và máy thứ hai tương ứng. Rõ ràng ta có một nhóm đầy đủ các biến cố xung khắc. Gọi  $B$  là biến cố lấy ra được một phế phẩm.

Ta có:

$$P(A_1) = \frac{2}{3} \quad , \quad P_{A_1}(B) = 0,03$$

$$P(A_2) = \frac{1}{3} \quad , \quad P_{A_2}(B) = 0,02$$

Theo công thức xác suất toàn phần:

$$P(B) = \sum_{i=1}^2 P(A_i)P_{A_i}(B) = \frac{2}{3} \cdot 0,03 + \frac{1}{3} \cdot 0,02 = \frac{0,08}{3}$$

b) Vì biến cố lấy một sản phẩm là phế phẩm ( $B$ ) và không là phế phẩm ( $\bar{B}$ ) là đối lập, nên ta có:

$$P(\bar{B}) = 1 - P(B) = 1 - \frac{0,08}{3} = \frac{2,92}{3}$$

**Bài 2:** Trong số 18 xạ thủ, nhóm I gồm 5 người bắn trúng đích với xác suất 0,8; nhóm II gồm 7 người bắn trúng đích với xác suất 0,7; nhóm III gồm 4 người bắn trúng đích với xác suất 0,6 và nhóm IV gồm 2 người bắn trúng đích với xác suất 0,5. Chọn hù họa một xạ thủ và cho anh ta bắn một phát, nhưng kết quả không trúng bia; xạ thủ ấy có khả năng thuộc nhóm nào nhiều nhất.



**Giải:** Gọi  $A_1, A_2, A_3, A_4$  là biến cố chọn ra một xạ thủ thuộc nhóm I, II, III, IV tương ứng. Rõ ràng ta có một nhóm đầy đủ các biến cố xung khắc. Gọi B là biến cố xạ thủ chọn ra bắn không trúng bia. Ta có:

$$P(A_1) = \frac{5}{18} \quad , \quad P_{A_1}(B) = 1 - 0,8 = 0,2$$

$$P(A_2) = \frac{7}{18} \quad , \quad P_{A_2}(B) = 1 - 0,7 = 0,3$$

$$P(A_3) = \frac{4}{18} \quad , \quad P_{A_3}(B) = 1 - 0,6 = 0,4$$

$$P(A_4) = \frac{2}{18} \quad , \quad P_{A_4}(B) = 1 - 0,5 = 0,5$$

Theo công thức Bâyét, ta có :

$$P_B(A_1) = \frac{P(A_1)P_{A_1}(B)}{\sum_{i=1}^4 P(A_i)P_{A_i}(B)} = \frac{1/18}{5,7/18} = \frac{1}{5,7}$$

$$P_B(A_2) = \frac{P(A_2)P_{A_2}(B)}{\sum_{i=1}^4 P(A_i)P_{A_i}(B)} = \frac{2,1}{5,7}$$

$$P_B(A_3) = \frac{P(A_3)P_{A_3}(B)}{\sum_{i=1}^4 P(A_i)P_{A_i}(B)} = \frac{1,6}{5,7}$$

$$P_B(A_4) = \frac{P(A_4)P_{A_4}(B)}{\sum_{i=1}^4 P(A_i)P_{A_i}(B)} = \frac{1}{5,7}$$

Suy ra :  $P_B(A_2) > P_B(A_3) > P_B(A_1) = P_B(A_4)$ , nghĩa là xạ thủ chọn ra bắn không trúng bia có khả năng thuộc nhóm II nhiều nhất.

**Bài 3:** Một xí nghiệp có 2 phân xưởng với các tỷ lệ phế phẩm tương ứng là 1% và 2%. Biết rằng phân xưởng I sản xuất 40%, còn phân xưởng II sản xuất 60% sản phẩm.

a) Tìm xác suất để từ kho xí nghiệp chọn ngẫu nhiên được một phế phẩm.

b) Giả sử lấy được 1 phế phẩm, tìm xác suất để nó do phân xưởng I sản xuất.

**Giải:** Gọi  $A_1$  và  $A_2$  là biến cố lấy ra 1 sản phẩm của phân xưởng I và II tương ứng, rõ ràng ta có một nhóm đầy đủ các biến cố. Gọi B là biến cố lấy ra được 1 phế phẩm.

a) Theo công thức xác suất đầy đủ:

$$P(B) = P(A_1)P_{A_1}(B) + P(A_2)P_{A_2}(B)$$

$$= 0,4.1\% + 0,6.2\% = 1,6\%.$$

Đó cũng chính là tỷ lệ phế phẩm chung của xí nghiệp.

b) Biến cố cần tính xác suất là  $A_1$  với điều kiện đã xảy ra B. Theo công thức Bêýt.

$$P_B(A_1) = \frac{P(A_1)P_{A_1}(B)}{\sum_{i=1}^2 P(A_i)P_{A_i}(B)} = \frac{0,4\%}{1,6\%} = \frac{1}{4}$$

**Bài 4:** Có 3 hộp bi giống nhau: hộp I chứa 20 bi trắng; hộp II chứa 10 bi trắng và 10 bi đen; còn hộp III chứa 20 bi đen. Chọn hú hoặ ra một hộp và từ đó rút hú hoặ ra được viên bi trắng. Tìm xác suất đó là viên bi của hộp thứ nhất.

**Giải:** Nhóm đầy đủ gồm 3 biến cố  $A_i$  ( $i=1, 2, 3$ ), ký hiệu cho việc chọn các hộp thứ  $i$  tương ứng. Dễ thấy  $P(A_1) = P(A_2) = P(A_3) = 1/3$ . Gọi B là biến cố rút được bi trắng và ta có ngay  $P_{A_1}(B) = 1$  (xác suất để rút được bi trắng từ hộp I);  $P_{A_2}(B) = 1/2$ ;  $P_{A_3}(B) = 0$ . Theo công thức Bêýt:

$$P_B(A_1) = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}$$

**Bài 5:** Một trạm chỉ phát hai loại tín hiệu A và B với xác suất tương ứng 0,84 và 0,16. Do có nhiễu trên đường truyền nên 1/6 tín



hiệu A bị méo và được thu như là tín hiệu B, còn 1/8 tín hiệu B bị méo thành tín hiệu A.

a) Tìm xác suất thu được tín hiệu A.

b) Giả sử thu được tín hiệu A, tìm xác suất để thu được đúng tín hiệu lúc phát.

**Giải:** a) Gọi  $H_A$  và  $H_B$  là biến cố tín hiệu A và B tương ứng đã được phát, ta có  $P(H_A)=0,84$ ;  $P(H_B)=0,16$ ; và chúng tạo nên nhóm đầy đủ. Gọi luôn A là biến cố thu được tín hiệu A. Theo điều kiện đầu bài:

$$P_{H_A}(A) = 1 - \frac{1}{6} = \frac{5}{6} ; P_{H_B}(B) = \frac{1}{8}$$

Theo công thức xác suất đầy đủ

$$\begin{aligned} P(A) &= P(H_A)P_{H_A}(A) + P(H_B)P_{H_B}(B) \\ &= 0,84 \cdot \frac{5}{6} + 0,16 \cdot \frac{1}{8} = 0,72 \end{aligned}$$

b) Xác suất cần tìm chính là  $P_A(H_A)$ , theo Bayét:

$$P_A(H_A) = \frac{0,84 \cdot \frac{5}{6}}{0,72} = \frac{35}{36} \approx 0,97.$$

**Bài 6:** Một dây chuyền gồm 2 bộ phận nối tiếp, với xác suất làm việc tốt trong một khoảng thời gian nào đó của mỗi bộ phận tương ứng là  $p_1$  và  $p_2$ . Ở một thời điểm trong khoảng thời gian trên người ta thấy dây chuyền hỏng (giả sử việc hỏng xảy ra chỉ do các bộ phận không làm việc). Hãy tìm xác suất để chỉ có bộ phận thứ nhất không làm việc.

**Giải:** Do các bộ phận mắc nối tiếp nên chỉ cần một bộ phận dừng là dây chuyền hỏng. Có thể xảy ra 4 khả năng khác nhau  $A_0$ -cả 2 bộ phận tốt;  $A_1$ - bộ phận I hỏng, bộ phận II tốt;  $A_2$ -bộ phận II hỏng còn bộ phận I tốt và  $A_3$ -cả 2 bộ phận hỏng.

$$\text{Dễ thấy: } P(A_0) = p_1 p_2 ; \quad P(A_1) = (1 - p_1) p_2 ;$$

$$P(A_2) = p_1 (1 - p_2) ; \quad P(A_3) = (1 - p_1)(1 - p_2).$$

Biến cố dây chuyền hỏng, ký hiệu là B, có:

$$P_{A_0}(B) = 0; \quad P_{A_1}(B) = P_{A_2}(B) = P_{A_3}(B) = 1$$

Theo công thức Bâyét:

$$\begin{aligned} P_B(A_1) &= \frac{(1 - p_1)p_2}{(1 - p_1)p_2 + p_1(1 - p_2) + (1 - p_1)(1 - p_2)} \\ &= \frac{(1 - p_1)p_2}{1 - p_1p_2} \end{aligned}$$

**Bài 7:** Cặp sinh đôi được gọi là thật nếu do cùng một trứng sinh ra và trong trường hợp này bao giờ cũng cùng giới tính. Nếu cặp đó do các trứng khác nhau sinh ra thì xác suất để cặp cùng giới tính là  $1/2$ . Nếu biết một cặp trẻ sinh đôi có cùng giới tính thì xác suất chúng là cặp sinh đôi thật là bao nhiêu, biết rằng xác suất để cặp sinh đôi do cùng một trứng sinh ra bằng  $p$ ?

**Giải:** Gọi  $A_1$ -biến cố cặp sinh đôi thật (cùng do một trứng sinh ra) thì  $A_2 = \bar{A}_1$  là biến cố đối lập (do hai trứng sinh ra)

$$P(A_1) = p; \quad P(A_2) = 1 - p.$$

Gọi B là biến cố cặp sinh đôi có cùng giới tính, dễ thấy:

$$P_{A_1}(B) = 1 \quad ; \quad P_{A_2}(B) = \frac{1}{2}$$

Từ đó theo công thức Bâyét:

$$P_B(A_1) = \frac{1 \cdot p}{1 \cdot p + \frac{1}{2}(1 - p)} = \frac{2p}{p + 1}$$

**Bài 8:** Một thiết bị có 10 chi tiết với độ tin cậy (xác suất làm việc tốt trong một khoảng thời gian nào đó) của mỗi chi tiết là 0,9. Tìm xác suất để trong khoảng thời gian ấy:

- a) Có đúng một chi tiết làm việc tốt;
- b) Có ít nhất 2 chi tiết làm việc tốt.

**Giải:** Dùng công thức Bécnuili với  $n = 10$ ,  $p = 0,9$ .

- a) Xác suất cần tìm là  $P_{10}(1) = C_{10}^1 0,9^1 0,1^9 = 0,9 \cdot 10^{-8}$ .



b) Xác suất cần tìm là:

$$1 - (P_{10}(0) + P_{10}(1)) = 1 - 0,1^{10} - C_{10}^1 0,9^1 0,1^9 = 1 - 0,91 \cdot 10^{-8}$$

**Bài 9:** Gieo 5 lần một đồng tiền cân đối đồng chất. Tìm xác suất xuất hiện:

- a) Đúng 1 lần mặt sấp;
- b) Hai lần mặt sấp;
- c) Ít nhất 1 lần mặt sấp.

**Giải:** Xác suất để xuất hiện mặt sấp (hoặc mặt ngửa) bằng  $1/2$ . Theo công thức Bécnuili ta có:

$$a) P_5(1) = C_5^1 \left(\frac{1}{2}\right)^5 = \frac{5}{32}$$

$$b) P_5(2) = C_5^2 \left(\frac{1}{2}\right)^5 = \frac{5}{16}$$

c) Ta tính qua sự kiện đối lập là không có xuất hiện mặt sấp tức là cả 5 lần đều mặt ngửa, từ đó xác suất cần tìm tính theo công thức Bécnuili bằng (chú ý  $0! = 1$  theo quy ước):

$$1 - C_5^0 \left(\frac{1}{2}\right)^5 = 1 - \frac{1}{32} = 31/32$$

**Bài 10:** Tỷ lệ phế phẩm của một lô hàng là 1%. Hỏi mẫu cần chọn ra là bao nhiêu (chọn có hoàn lại) sao cho xác suất  $> 0,95$  trong mẫu đó có ít nhất một phế phẩm?

**Giải:** Giả sử mẫu chọn ra có kích cỡ là  $n$  và việc chọn ra 1 sản phẩm có hoàn lại là một phép thử Bécnuili độc lập với  $p=0,01$ . Xác suất để trong mẫu có ít nhất một phế phẩm sẽ là:

$$1 - (1-p)^n = 1 - 0,99^n$$

Theo đầu bài yêu cầu:

$$1 - 0,99^n \geq 0,95 \Leftrightarrow 0,05 \geq 0,99^n$$

$$\Rightarrow \ln 0,05 \geq n \ln 0,99$$

$$\Rightarrow n \geq \frac{\ln 0,05}{\ln 0,99} > 296.$$

**Bài 11:** Một bác sỹ chữa bệnh có xác suất chữa khỏi là 0,8. Có người nói rằng cứ 5 người đến chữa thì có chắc chắn 4 người khỏi bệnh, người khác lại cho rằng trong 10 người đến chữa có chắc chắn 8 người khỏi bệnh. Điều đó có đúng không?

**Giải:** Cả hai người khẳng định đều sai. Xác suất xảy ra trường hợp thứ nhất là:

$$P_5(4) = C_5^4 \cdot 0,8^4 \cdot 0,2^1 = 0,4096$$

Còn xác suất xảy ra trường hợp thứ hai:

$$P_{10}(8) = C_{10}^8 \cdot 0,8^8 \cdot 0,2^2 \approx 0,3018$$

**Bài 12:** Một cầu thủ ném bóng 400 lần vào rổ với xác suất ném trúng rổ của mỗi lần ném là 0,8. Tìm xác suất để cầu thủ đó ném trúng 300 lần

**Giải:** Rõ ràng có thể dùng công thức Béc-nu-li:

$$P_{400}(300) = C_{400}^{300} (0,8)^{300} (0,2)^{100}$$

Tuy nhiên việc tính toán khá phức tạp. Ta sẽ dùng định lý giới hạn Laplat 1:

$$P_{400}(300) \approx \frac{1}{\sqrt{400 \cdot 0,8 \cdot 0,2}} \varphi(x_0)$$

với  $\varphi(x_0)$  là hàm số Gauxơ tại  $x_0 = \frac{k - np}{\sqrt{npq}}$ , tức ta có:

$$\varphi(x_0) = \varphi\left(\frac{300 - 400 \cdot 0,8}{\sqrt{400 \cdot 0,8 \cdot 0,2}}\right) = \varphi(-25) \approx 0,0175$$

Từ đó xác suất cần tìm  $P_{400}(300) \approx 0,0023$ .

**Bài 13:** Xác suất bắn trúng đích của một xạ thủ là 0,8. Tìm xác suất để trong 100 lần bắn

a) Xạ thủ bắn trúng không ít hơn 75 lần và không nhiều hơn 90 lần;

b) Không ít hơn 75 lần bắn trúng.



**Giải:** Dùng định lý giới hạn Laplat 2:

$$P_n(k_1; k_2) = \Phi(x_2) - \Phi(x_1),$$

$$\text{với: } x_i = \frac{k_i - np}{\sqrt{npq}} \quad (i=1,2)$$

và hàm số Laplat  $\Phi(x_i)$ .

$$\text{a) } P_{100}(75; 90) = \Phi(x_2) - \Phi(x_1),$$

$$\text{với: } x_1 = \frac{75 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = -1,25; \quad x_2 = \frac{90 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = 2,5$$

$$\text{Từ đó: } P_{100}(75;90) = \Phi(2,5) - \Phi(-1,25) \approx 0,8882.$$

b) Tương tự ta có:

$$P_{100}(75;100) = \Phi(5) - \Phi(-1,25) \approx 0,8943.$$

**Bài 13:** Một nữ công nhân đứng máy xe sợi gồm 800 ống sợi, xác suất đứt sợi của mỗi ống trong vòng 1 giờ là 0,005. Tìm xác suất của sự kiện trong vòng 1 giờ có 4 ống sợi bị đứt.

**Giải:** Ở đây  $p=0,005$  rất bé nên không dùng xấp xỉ Laplat mà dùng xấp xỉ theo Poátxông với tham số  $\lambda = np = 800 \cdot 0,005 = 4$ .

$$P_{800}(4) \approx \frac{4^4 \cdot e^{-4}}{4!} = 0,1954.$$

### Chương 3

## ĐẠI LƯỢNG NGẪU NHIÊN. CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

### 3.1. KHÁI NIỆM VỀ ĐẠI LƯỢNG NGẪU NHIÊN. BẢNG PHÂN PHỐI XÁC SUẤT. HÀM PHÂN PHỐI XÁC SUẤT

#### 3.1.1. Đại lượng ngẫu nhiên

*a) Bài toán dẫn đến khái niệm:* Gieo một con xúc xắc xuống mặt bàn, kết quả phép thử đó là một số, số này biến đổi từ 1 đến 6 với xác suất xác định tương ứng đều bằng  $1/6$ .

Từ bài toán trên ta có nhận xét: Khi thực hiện một phép thử thì kết quả của phép thử ta có thể biểu diễn được bằng một đại lượng biến đổi, đại lượng này chỉ có thể nhận một trong các giá trị nào đó với xác suất tương ứng xác định. Một đại lượng biến đổi như vậy cho ta một khái niệm gọi là đại lượng ngẫu nhiên trong lý thuyết xác suất.

*b) Định nghĩa:* Đại lượng ngẫu nhiên là một đại lượng mà trong kết quả phép thử nó chỉ có thể nhận một trong các giá trị có thể nào đó với một xác suất tương ứng xác định.

Ta ký hiệu đại lượng ngẫu nhiên là những chữ cái in hoa:  $X_i, Y_i, Z_i, \dots$  và các giá trị có thể của chúng là  $x_i, y_i, z_i$  ( $i=1, 2, 3, \dots, n$ ).

*c) Các loại đại lượng ngẫu nhiên:* Có hai loại đại lượng ngẫu nhiên, đại lượng ngẫu nhiên liên tục và đại lượng ngẫu nhiên rời rạc.



Đại lượng ngẫu nhiên là rời rạc nếu các giá trị có thể của nó lập thành một chuỗi số hữu hạn hay vô hạn. Ví dụ: Đại lượng ngẫu nhiên khi tung hạt xúc xắc (1, 2, 3, 4, 5, 6) là đại lượng ngẫu nhiên rời rạc.

Đại lượng ngẫu nhiên là liên tục nếu các giá trị có thể của nó lấp kín một khoảng hữu hạn hay vô hạn. Ví dụ: Đại lượng ngẫu nhiên chỉ chiều cao của tập hợp thanh niên từ 15 đến 20 tuổi là đại lượng ngẫu nhiên liên tục.

### 3.1.2. Bảng phân phối xác suất

Về nguyên tắc một đại lượng ngẫu nhiên hoàn toàn được xác định khi ta có tất cả các giá trị có thể của nó và các xác suất tương ứng. Gọi  $p_i$  là xác suất của biến cố  $x_i$  ( $i=1, 2, \dots, n$ ) lúc đó đại lượng ngẫu nhiên  $x$  gồm  $x_1, x_2, \dots, x_n$  được xác định khi ta có xác suất tương ứng  $p$  gồm  $p_1, p_2, \dots, p_n$ . Trong trường hợp đại lượng ngẫu nhiên là rời rạc, người ta thường lập bảng dạng:

$x$	$x_1$	$x_2$	.....	$x_n$
$p$	$p_1$	$p_2$	.....	$p_n$

Bảng này được gọi là bảng phân phối xác suất dùng để xác định đại lượng ngẫu nhiên.

*Ví dụ:* Bảng phân phối xác suất của đại lượng ngẫu nhiên chỉ số chấm khi tung hạt xúc xắc là bảng sau:

$x$	1	2	3	4	5	6
$p$	1/6	1/6	1/6	1/6	1/6	1/6

### 3.1.3. Hàm phân phối xác suất

Khi đại lượng ngẫu nhiên là liên tục ta không thể viết tất cả các giá trị có thể của nó cùng các xác suất tương ứng và lập bảng phân phối xác suất được. Trong trường hợp này ta xác định đại lượng ngẫu

nhiên bằng cách khác trên cơ sở khái niệm về hàm phân phối xác suất của đại lượng ngẫu nhiên.

**a) Định nghĩa:** Hàm phân phối xác suất của một đại lượng ngẫu nhiên  $x$  được cho bằng biểu thức:

$$F(x) = P(X < x) \quad (1)$$

đó là hàm đối số  $x$  để xác định xác suất khi đại lượng ngẫu nhiên  $X$  nhận các giá trị nhỏ hơn  $x$  (theo định nghĩa này hàm phân phối xác suất dùng cho cả đại lượng ngẫu nhiên rời rạc và liên tục).

**b) Ví dụ:** Xây dựng hàm phân phối xác suất của đại lượng ngẫu nhiên chỉ số chấm khi tung hạt xúc xắc.

$$\text{Khi } x \leq 1 \text{ thì } F(x) = P(X < x) \equiv P(X < 1) = 0$$

$$\text{Khi } 1 < x \leq 2 \text{ thì } F(x) = P(X < x) \equiv P(X < 2) = P(x_1=1) = 1/6$$

$$\begin{aligned} \text{Khi } 2 < x \leq 3 \text{ thì } F(x) = P(X < x) &\equiv P(X < 3) = P(x_1=1) + P(x_2=2) \\ &= \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} \text{Khi } 3 < x \leq 4 \text{ thì } F(x) &= P(X < x) \equiv P(X < 4) = P(x_1=1) + P(x_2=2) + P(x_3=3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \text{Khi } 4 < x \leq 5 \text{ thì } F(x) &= P(X < x) \equiv P(X < 5) \\ &= P(x_1=1) + P(x_2=2) + P(x_3=3) + P(x_4=4) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3} \end{aligned}$$

$$\begin{aligned} \text{Khi } 5 < x \leq 6 \text{ thì } F(x) &= P(X < x) \equiv P(X < 6) \\ &= P(x_1=1) + P(x_2=2) + P(x_3=3) + P(x_4=4) + P(x_5=5) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{5}{6} \end{aligned}$$

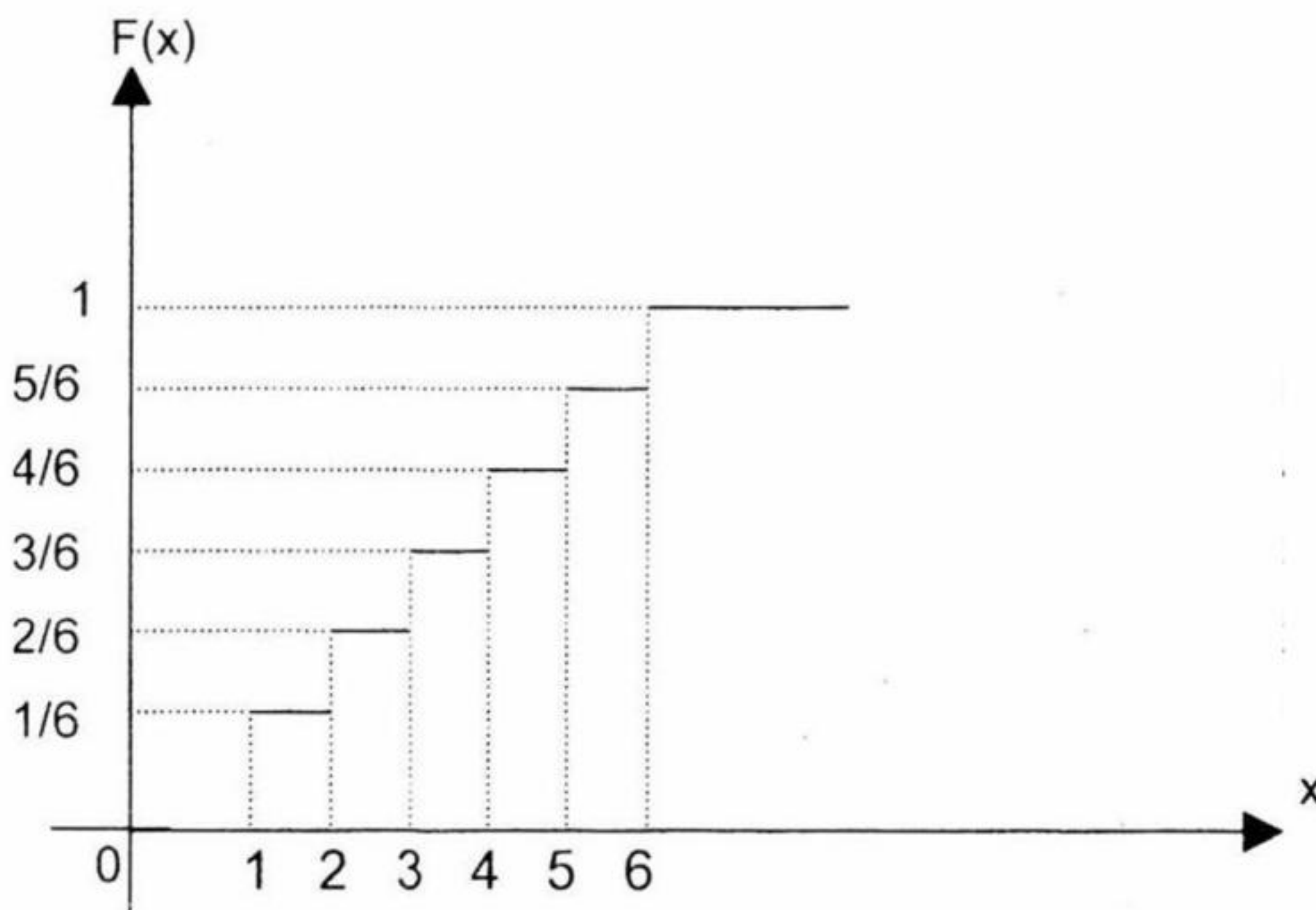
$$\text{Khi } x > 6 \text{ thì } F(x) = P(x=6) = 1.$$

Tóm lại ta có hàm phân phối xác suất:



$$F(x) = \begin{cases} 0 & , \quad x \leq 1 \\ 1/6 & , \quad 1 < x \leq 2 \\ 1/3 & , \quad 2 < x \leq 3 \\ 1/2 & , \quad 3 < x \leq 4 \\ 2/3 & , \quad 4 < x \leq 5 \\ 5/6 & , \quad 5 < x \leq 6 \\ 1 & , \quad x > 6 \end{cases} \quad (2)$$

Đồ thị của hàm  $F(x)$  có dạng hình 12.



**Hình 12**

c) Tính chất của hàm phân phối xác suất:

**\* Tính chất 1:**

Hàm phân phối xác suất  $F(x)$  không âm, có giá trị giữa 0 và 1, nghĩa là:

$$0 \leq F(x) \leq 1 \quad (3)$$

*Chứng minh:* Tính chất này được suy ra từ định nghĩa hàm  $F(x)$  là xác suất  $P(X < x)$  và tính chất của xác suất một biến cố.

**\* Tính chất 2:**

Hàm phân phối xác suất là hàm không giảm, nghĩa là với  $x_2 > x_1$  thì

$$F(x_2) \geq F(x_1) \quad (4)$$

*Chứng minh:* Biến cố  $(X < x_2)$  là tổng của hai biến cố xung khắc  $(X < x_1)$  và  $(x_1 \leq X < x_2)$ , nghĩa là ta có:

$$(X < x_2) = (X < x_1) + (x_1 \leq X < x_2) \quad (5)$$

Tương ứng ta có xác suất của tổng hai biến cố xung khắc:

$$P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2) \quad (6)$$

Theo định nghĩa hàm phân phối xác suất:

$$\begin{aligned} P(X < x_1) &= F(x_1) \\ P(X < x_2) &= F(x_2) \end{aligned} \quad (7)$$

Từ (6) và (7), ta có :

$$F(x_2) - F(x_1) = P(x_1 \leq X < x_2) \quad (8)$$

Mặt khác, theo tính chất xác suất của biến cố:

$$0 \leq P(x_1 \leq X < x_2) \leq 1 \quad (9)$$

Từ (8) và (9) suy ra:

$$F(x_2) - F(x_1) \geq 0 \quad (10)$$

Đây chính là điều cần chứng minh.

**\* Tính chất 3:**

Nếu đại lượng ngẫu nhiên  $X$  chỉ nhận các giá trị trong khoảng đóng  $[a, b]$  thì hàm phân phối xác suất  $F(x)$  có tính chất:

$$\begin{aligned} F(x) &= 0 \quad \text{nếu} \quad x \leq a \\ F(x) &= 1 \quad \text{nếu} \quad x > b \end{aligned} \quad (11)$$



*Chứng minh:*

Nếu  $x \leq a$ , thì ta có:

$$F(x) = P(X < x) \equiv P(X < a) = 0$$

vì  $P(X < a)$  là xác suất của biến cố không thể có và bằng 0.

Nếu  $x > b$ , thì ta có:

$$F(x) = P(X < x) \equiv P(X \leq b) = 1$$

vì  $P(X \leq b)$  là xác suất của biến cố chắc chắn và bằng 1.

### **3.1.4. Hàm mật độ phân phối xác suất**

*a) Định nghĩa:*

Nếu hàm phân phối xác suất của đại lượng ngẫu nhiên  $X$  có đạo hàm bậc nhất thì đạo hàm bậc nhất đó được gọi là hàm mật độ phân phối xác suất của đại lượng ngẫu nhiên  $X$  và ký hiệu là  $f(x)$ , nghĩa là:

$$f(x) = \frac{dF(x)}{dx} \quad (12)$$

*b) Tính chất:*

**\* Tính chất 1:**

Hàm mật độ phân phối xác suất là hàm không âm, nghĩa là:

$$f(x) \geq 0 \quad (13)$$

*Chứng minh:*

Vì  $f(x)$  là đạo hàm của hàm không giảm  $F(x)$  nên theo tính chất của đạo hàm thì  $f(x) \geq 0$ .

**\* Tính chất 2:**

Hàm phân phối xác suất  $F(x)$  của đại lượng ngẫu nhiên bằng tích phân của hàm mật độ phân phối xác suất từ  $-\infty$  đến  $x$ , nghĩa là:

$$F(x) = \int_{-\infty}^x f(x') dx' \quad (14)$$

*Chứng minh:*

Từ định nghĩa hàm mật độ phân phối xác suất ta có hệ thức:

$$dF(x)=f(x) dx \quad (15)$$

Lấy tích phân hai vế của (15) từ  $-\infty$  đến  $x$ , ta có:

$$\int_{-\infty}^x f(x')dx' = \int_{-\infty}^x dF(x') = F(x) - F(-\infty) \quad (16)$$

Mặt khác theo định nghĩa hàm phân phối xác suất:

$$F(-\infty) = P(X < -\infty) = 0 \quad (17)$$

(Vì  $P(X < -\infty)$  là xác suất của biến cố không thể có và bằng 0).

Từ (16) và (17) suy ra:

$$\int_{-\infty}^x f(x')dx' = F(x)$$

Đây chính là điều 2 cần chứng minh.

**\* Tính chất 3:**

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad (18)$$

*Chứng minh:*

Theo tính chất 2 ta có:

$$F(x) = \int_{-\infty}^x f(x')dx'$$

cho  $x \rightarrow +\infty$ , ta có:

$$\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow +\infty} \int_{-\infty}^x f(x')dx'$$

$$F(+\infty) = \int_{-\infty}^{+\infty} f(x')dx'$$

Nhưng  $F(+\infty) = p(X < +\infty) = 1$ , (vì  $p(X < +\infty)$  là xác suất của biến cố chắc chắn và bằng 1).



Suy ra:

$$\int_{-\infty}^{+\infty} f(x') dx' = 1$$

Đây chính là điều cần chứng minh.

c) Ví dụ:

Hàm mật độ phân phối xác suất  $f(x)$  của đại lượng ngẫu nhiên  $X$  có dạng:

$$f(x) = \begin{cases} c & \text{khi } a \leq x \leq b \\ 0 & \text{khi } x < a \text{ hoặc } x > b \end{cases}$$

Hãy tìm hằng số  $c$ , hàm phân phối xác suất và vẽ đồ thị của hàm phân phối xác suất  $F(x)$ .

*Giải:* Theo tính chất 3 của hàm mật độ phân phối xác suất ta có:

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^b f(x) dx + \int_b^{+\infty} f(x) dx = \int_a^b c dx = c(b - a)$$

$$\text{Suy ra : } c = \frac{1}{b - a}.$$

Theo tính chất 2 của hàm mật độ phân phối xác suất ta có:

$$F(x) = \int_{-\infty}^x f(x') dx'$$

Nếu  $x < a$  thì  $f(x') = 0$ , dẫn đến  $F(x) = 0$ .

Nếu  $a \leq x \leq b$  thì:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(x') dx' = \int_{-\infty}^a f(x') dx' + \int_a^x f(x') dx' = \\ &= \int_a^x c dx' = c(x - a) = \frac{x - a}{b - a} \end{aligned}$$

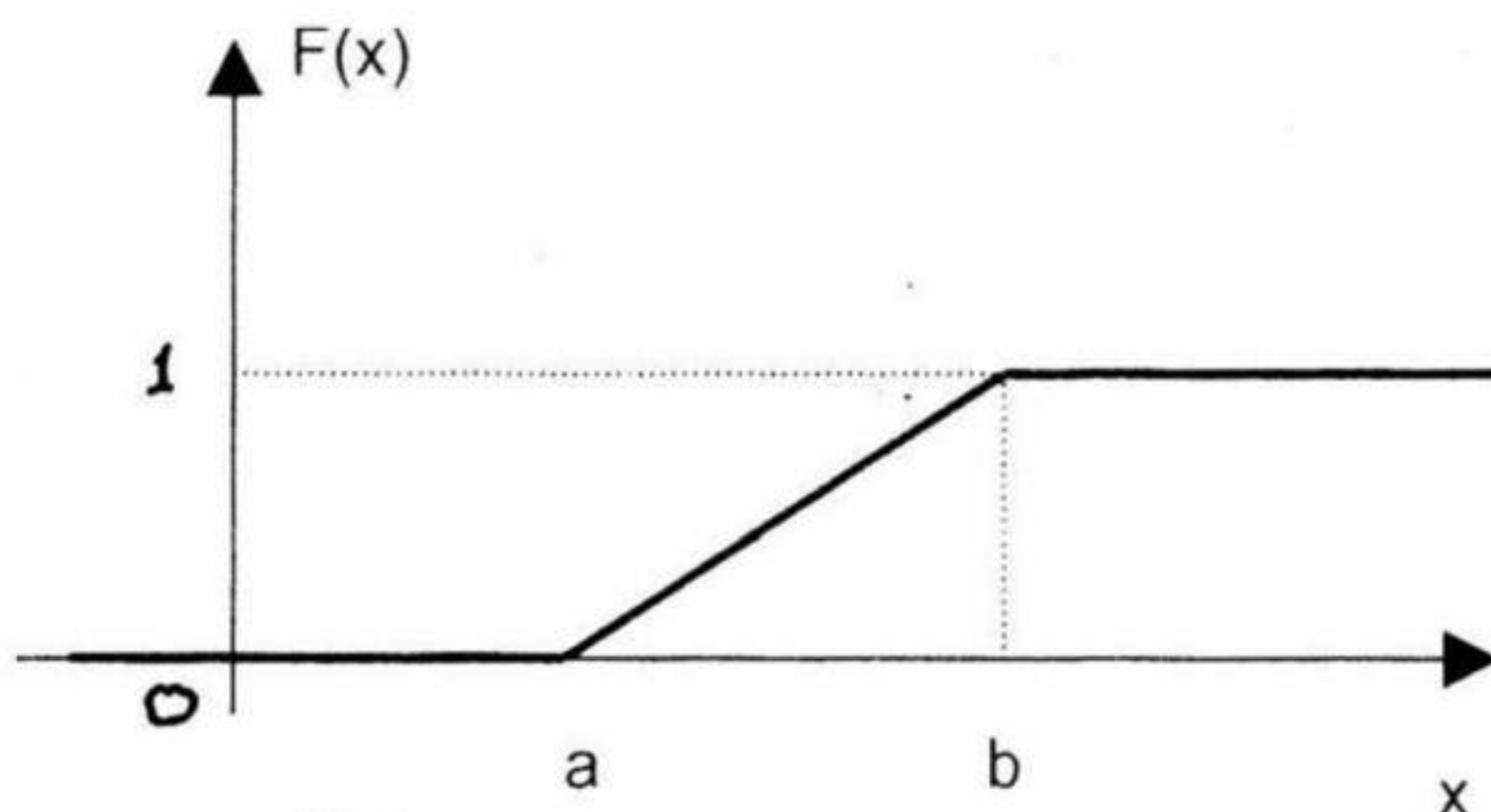
Nếu  $x > b$  thì:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(x') dx' = \int_{-\infty}^a f(x') dx' + \int_a^b f(x') dx' + \int_b^x f(x') dx' = \\ &= \int_a^b c dx' = c(b-a) = \frac{b-a}{b-a} = 1 \end{aligned}$$

Tóm lại, ta có:

$$F(x) = \begin{cases} 0 & , \quad x < a \\ \frac{x-a}{b-a} & , \quad a \leq x \leq b \\ 1 & , \quad x > b \end{cases}$$

Đồ thị hàm số  $F(x)$  có dạng:



## 3.2. KỲ VỌNG TOÁN CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Một tham số đặc trưng quan trọng của đại lượng ngẫu nhiên đó là kỳ vọng toán của đại lượng ngẫu nhiên.

### 3.2.1. Định nghĩa 1

Kỳ vọng toán của đại lượng ngẫu nhiên rời rạc  $X$  là tổng các tích giữa các giá trị có thể của nó với xác suất tương ứng xác định.



Ta ký hiệu kỳ vọng toán của đại lượng ngẫu nhiên  $X$  là  $E(X)$ . Khi đó theo định nghĩa 1 ta có biểu thức toán học của kỳ vọng toán  $E(X)$ :

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i \quad (19)$$

### 3.2.2. Định nghĩa 2

Kỳ vọng toán của đại lượng ngẫu nhiên liên tục  $X$  với miền giá trị là đoạn  $[a, b]$  (khoảng đóng  $[a, b]$ ), là một số xác định bởi tích phân:

$$E(X) = \int_a^b x f(x) dx \quad (20)$$

( $f(x)$  là hàm mật độ phân phối xác suất).

### 3.2.3. Các ví dụ về kỳ vọng toán của đại lượng ngẫu nhiên

*Ví dụ 1:* Tìm kỳ vọng toán của đại lượng ngẫu nhiên  $X$  chỉ số chấm khi tung hạt xúc xắc.

*Giải:* Theo công thức (19) ta có:

$$\begin{aligned} E(X) &= x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 + x_6 p_6 = \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3,5 \end{aligned}$$

*Ví dụ 2:* Tìm kỳ vọng toán của đại lượng ngẫu nhiên liên tục, phân phối đều trong khoảng  $[a, b]$  và có hàm mật độ phân phối xác suất  $f(x)$  dạng:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{khi } x \in [a, b] \\ 0 & \text{khi } x \notin [a, b] \end{cases}$$

*Giải:* Theo công thức (20), ta có:

$$E(X) = \int_a^b \frac{x}{(b-a)} dx = \frac{1}{(b-a)} \cdot \frac{x^2}{2} \Big|_a^b = \frac{1}{2}(b+a)$$

### 3.2.4. Tính chất của kỳ vọng toán

#### a) Tính chất 1

Kỳ vọng toán của một hằng số thì bằng chính hằng số đó, nghĩa là:

$$E(c) = c \quad (c - \text{hằng số}) \quad (21)$$

*Chứng minh:*

Vì có thể coi hằng số  $c$  như một biến cố chắc chắn với xác suất bằng 1, ta có theo định nghĩa kỳ vọng toán (19):

$$E(c) = c \cdot 1 = c$$

#### b) Tính chất 2

Thừa số không đổi có thể đưa ra ngoài ký hiệu kỳ vọng toán, nghĩa là:

$$E(cX) = cE(X) \quad (c - \text{hằng số}) \quad (22)$$

**Chứng minh:**

Nếu đại lượng ngẫu nhiên  $X$  với các giá trị có thể của nó là  $x_1, x_2, \dots, x_j, \dots$  và các xác suất tương ứng  $p_1, p_2, \dots, p_j, \dots$  thì đại lượng ngẫu nhiên  $cX$  với các giá trị có thể của nó là  $cx_1, cx_2, \dots, cx_j, \dots$  và cũng với các xác suất tương ứng  $p_1, p_2, \dots, p_j, \dots$ . Theo định nghĩa kỳ vọng toán ta có:

$$E(cX) = \sum_{i=1}^n cx_i p_i = c \sum_{i=1}^n x_i p_i = cE(X)$$

#### c) Tính chất 3

Kỳ vọng toán của tổng hai đại lượng ngẫu nhiên bằng tổng các kỳ vọng toán của chúng, nghĩa là:

$$E(X+Y) = E(X) + E(Y) \quad (23)$$

*Chứng minh:*

Không làm mất đi tính tổng quát khi chúng ta giả thiết đại lượng ngẫu nhiên  $X, Y$  là rời rạc và chỉ nhận 2 giá trị, nghĩa là ta có:



	$x_1$	$x_2$
$X$		
<hr/>		
$P$	$p_1$	$p_2$
<hr/>		
	$y_1$	$y_2$
$Y$		
<hr/>		
$G$	$g_1$	$g_2$
<hr/>		

Lúc đó, đại lượng ngẫu nhiên  $(X+Y)$  sẽ nhận các giá trị  $(x_1+y_1)$ ,  $(x_1+y_2)$ ,  $(x_2+y_1)$ ,  $(x_2+y_2)$  với các xác suất tương ứng  $q_{11}$ ,  $q_{12}$ ,  $q_{21}$ ,  $q_{22}$ . Nghĩa là, ta có theo định nghĩa kỳ vọng toán:

$$\begin{aligned} E(X+Y) &= (x_1+y_1)q_{11} + (x_1+y_2)q_{12} + (x_2+y_1)q_{21} + (x_2+y_2)q_{22} = \\ &= [x_1(q_{11}+q_{12}) + x_2(q_{21}+q_{22})] + [y_1(q_{11}+q_{21}) + y_2(q_{12}+q_{22})] \quad (24) \end{aligned}$$

Từ (24) suy ra tính chất 3 sẽ được chứng minh nếu ta chứng minh được các hệ thức:

$$q_{11} + q_{12} = p_1, \quad q_{21} + q_{22} = p_2$$

$$q_{11} + q_{21} = g_1, \quad q_{12} + q_{22} = g_2$$

Để làm điều đó, trước tiên ta chứng minh hệ thức  $q_{11} + q_{12} = p_1$ . Xét biến cố  $x_1$  của đại lượng ngẫu nhiên  $X$ . Lúc đó đại lượng ngẫu nhiên  $(X+Y)$  hoặc nhận giá trị  $(x_1+y_1)$  hoặc nhận giá trị  $(x_1+y_2)$ . Do đó, biến cố  $x_1$  sẽ là tổng của hai biến cố xung khắc  $(x_1+y_1)$  và  $(x_1+y_2)$ , nghĩa là:

$$x_1 = (x_1+y_1) + (x_1+y_2)$$

$$Px_1 = P(x_1+y_1) + P(x_1+y_2) \quad \text{hay} \quad p_1 = q_{11} + q_{12}$$

Các hệ thức khác cũng được chứng minh tương tự. Dẫn đến:

$$\begin{aligned} E(X + Y) &= [x_1 p_1 + x_2 p_2] + [y_1 g_1 + y_2 g_2] = \\ &= E(X) + E(Y) \end{aligned}$$

**d) Tính chất 4:**

Kỳ vọng toán của tích hai đại lượng ngẫu nhiên độc lập bằng tích các kỳ vọng toán của chúng, nghĩa là:

$$E(X.Y) = E(X).E(Y) \quad (25)$$

(với X, Y là hai đại lượng ngẫu nhiên độc lập).

**Chứng minh:**

Không làm mất đi tính tổng quát khi chúng ta giả thiết đại lượng ngẫu nhiên độc lập X, Y là rời rạc và chỉ nhận 2 giá trị như phần chứng minh tính chất 3. Nghĩa là:

X	$x_1$	$x_2$
P	$p_1$	$p_2$
Y	$y_1$	$y_2$
G	$g_1$	$g_2$

Lúc đó, đại lượng ngẫu nhiên độc lập X.Y sẽ nhận các giá trị  $x_1 y_1$ ,  $x_1 y_2$ ,  $x_2 y_1$ ,  $x_2 y_2$  với các xác suất tương ứng  $p_1 g_1$ ,  $p_1 g_2$ ,  $p_2 g_1$ ,  $p_2 g_2$  (vì ở đây ta có xác suất của tích hai biến cố độc lập bằng tích các xác suất của chúng). Nghĩa là, theo định nghĩa kỳ vọng toán của đại lượng ngẫu nhiên ta có:

$$\begin{aligned} E(X.Y) &= x_1 y_1 p_1 g_1 + x_1 y_2 p_1 g_2 + x_2 y_1 p_2 g_1 + x_2 y_2 p_2 g_2 = \\ &= x_1 p_1 (y_1 g_1 + y_1 g_2) + x_2 p_2 (y_1 g_1 + y_2 g_2) = \\ &= (x_1 p_1 + x_2 p_2) \cdot (y_1 g_1 + y_2 g_2) = E(X).E(Y) \end{aligned}$$



### 3.3. PHƯƠNG SAI CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Phương sai của đại lượng ngẫu nhiên phản ánh độ phân tán giữa đại lượng ngẫu nhiên với kỳ vọng toán của đại lượng ngẫu nhiên đó.

#### 3.3.1. Định nghĩa

Phương sai của đại lượng ngẫu nhiên  $X$  là kỳ vọng toán của bình phương sai lệch giữa đại lượng ngẫu nhiên với kỳ vọng toán của đại lượng ngẫu nhiên đó và ký hiệu là  $D(X)$ , nghĩa là:

$$D(X) = E[(X - m)^2] \quad (26)$$

(với  $m = E(X)$ ).

Nếu  $X$  là đại lượng ngẫu nhiên rời rạc với các giá trị có thể  $x_1, x_2, \dots, x_n$  và các xác suất tương ứng  $p_1, p_2, \dots, p_n$  thì phương sai của đại lượng ngẫu nhiên  $D(X)$  có dạng:

$$D(X) = \sum_{i=1}^n (x_i - m)^2 p_i \quad (27)$$

Nếu  $X$  là đại lượng ngẫu nhiên liên tục với miền giá trị là đoạn  $[a, b]$  (khoảng đóng  $[a, b]$ ), thì phương sai của đại lượng ngẫu nhiên  $D(X)$  có dạng:

$$D(X) = \int_a^b (x - m)^2 f(x) dx \quad (28)$$

( $f(x)$  là hàm mật độ phân phối xác suất).

#### 3.3.2. Các ví dụ về phương sai của đại lượng ngẫu nhiên

*Ví dụ 1:* Tìm phương sai của đại lượng ngẫu nhiên rời rạc  $x$  chỉ số chấm khi tung hạt xúc xắc.

*Giải:*

$$m = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

$$D(X) = E[(X - m)^2] = (1 - 3,5)^2 \cdot \frac{1}{6} + (2 - 3,5)^2 \cdot \frac{1}{6} + (3 - 3,5)^2 \cdot \frac{1}{6} + (4 - 3,5)^2 \cdot \frac{1}{6} + (5 - 3,5)^2 \cdot \frac{1}{6} + (6 - 3,5)^2 \cdot \frac{1}{6} = \frac{35}{12}$$

*Ví dụ 2:* Tìm phương sai của đại lượng ngẫu nhiên liên tục phân phối đều trong khoảng  $[a, b]$  và có hàm mật độ phân phối xác suất  $f(x)$  có dạng:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{khi } x \in [a, b] \\ 0 & \text{khi } x \notin [a, b] \end{cases}$$

*Giải:*

$$m = E(X) = \int_a^b \frac{x}{(b-a)} dx = \frac{1}{2}(b+a)$$

$$\begin{aligned} D(X) &= E[(X - m)^2] = \int_a^b \frac{(x - m)^2}{(b-a)} dx = \\ &= \frac{1}{(b-a)} \left[ \int_a^b x^2 dx - 2m \int_a^b x dx + m^2 \int_a^b dx \right] = \\ &= \frac{1}{(b-a)} \left[ \frac{(b^3 - a^3)}{3} - 2m \cdot \frac{(b^2 - a^2)}{2} + m^2(b-a) \right] = \\ &= \frac{1}{(b-a)} \left[ \frac{1}{3}(b^3 - a^3) - \frac{1}{2}(b-a)(b^2 - a^2) + \frac{1}{4}(b+a)^2(b-a) \right] \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

### 3.3.3. Dạng khác của công thức tính phương sai

Từ định nghĩa phương sai theo công thức (26) và các tính chất của kỳ vọng toán ta có thể viết công thức tính phương sai dưới dạng khác.

$$\begin{aligned} D(X) &= E[(X-m)^2] = E(X^2 - 2mX + m^2) = E(X^2) - E(2mX) + E(m^2) \\ &= E(X^2) - m^2, \text{ (với } m = E(X)) \end{aligned}$$



nghĩa là, ngoài công thức:

$$D(X) = E[(X-m)^2], \text{ với } m = E(X)$$

còn có thể tính  $D(X)$  theo công thức:

$$D(X) = E(X^2) - m^2, \text{ với } m = E(X) \quad (29)$$

Theo (29), tương tự như (27) trong trường hợp đại lượng ngẫu nhiên rời rạc ta có:

$$D(X) = \sum_{i=1}^n x_i^2 p_i - m^2 \quad (30)$$

và tương tự như (28) trong trường hợp đại lượng ngẫu nhiên liên tục ta có:

$$D(X) = \int_a^b x^2 f(x) dx - m^2 \quad (31)$$

### 3.3.4. Tính chất của phương sai

#### a) Tính chất 1:

Phương sai của một hằng số thì bằng không, nghĩa là:

$$D(c) = 0, \quad (c - \text{hằng số}) \quad (32)$$

*Chứng minh:*

Theo công thức (29), ta có:

$$D(c) = E(c^2) - [E(c)]^2 = c^2 - [c]^2 = 0$$

#### b) Tính chất 2:

Thừa số là hằng số không đổi có thể đưa ra ngoài dấu phương sai sau khi đã bình phương, nghĩa là:

$$D(cX) = c^2 D(X), \quad (c - \text{hằng số}) \quad (33)$$

*Chứng minh:*

Từ công thức (29), ta có:

$$\begin{aligned} D(cX) &= E(c^2 X^2) - [E(cX)]^2 = c^2 E(X^2) - c^2 [E(X)]^2 \\ &= c^2 \{E(X^2) - [E(X)]^2\} = c^2 D(X) \end{aligned}$$

**c) Tính chất 3:**

Phương sai của tổng hai đại lượng ngẫu nhiên độc lập bằng tổng các phương sai của chúng, nghĩa là:

$$D(X+Y) = D(X)+D(Y) \quad (34)$$

(với X, Y là hai đại lượng ngẫu nhiên độc lập).

**Chứng minh:**

Theo công thức (29), ta có:

$$\begin{aligned} D(X + Y) &= E[(X + Y)^2] - [E(X + Y)]^2 = \\ &= E(X^2) + E(2XY) + E(Y^2) - [E(X) + E(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \end{aligned} \quad (35)$$

Vì X, Y là hai đại lượng ngẫu nhiên độc lập nên theo tính chất của kỳ vọng toán của 2 đại lượng ngẫu nhiên độc lập  $E(XY) = E(X).E(Y)$ , nên từ (35) ta có:

$$\begin{aligned} D(X+Y) &= E(X^2) - 2E(X).E(Y) + E(Y^2) - [E(X)]^2 - 2E(X).E(Y) - [E(Y)]^2 \\ &= \{E(X^2) - [E(X)]^2\} + \{E(Y^2) - [E(Y)]^2\} = D(X) + D(Y) \end{aligned}$$

**3.3.5. Độ lệch quân phương của đại lượng ngẫu nhiên**

Để đánh giá mức độ phân tán giữa đại lượng ngẫu nhiên và kỳ vọng toán của nó đôi khi người ta sử dụng một tham số đặc trưng khác của đại lượng ngẫu nhiên gọi là độ lệch quân phương.

**a) Định nghĩa:**

Độ lệch quân phương của đại lượng ngẫu nhiên là căn bậc hai của phương sai của đại lượng ngẫu nhiên đó và ký hiệu là  $\sigma(X)$ , nghĩa là:

$$\sigma(X) = \sqrt{D(X)} \quad (36)$$

Theo định nghĩa này, tương tự như phương sai ta có các tính chất sau của độ lệch quân phương:

$$\sigma(c) = 0, \quad (c - \text{hằng số})$$



$$\sigma(cX) = c\sigma(X) \quad (c - \text{hằng số})$$

$$\sigma(X + Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}$$

(với  $X, Y$  là 2 đại lượng ngẫu nhiên độc lập).

*b) Ví dụ:*

Độ lệch quân phương của đại lượng ngẫu nhiên rời rạc  $X$  chỉ số chấm khi tung hạt xúc xắc bằng  $\sqrt{D(X)} = \left[ \frac{35}{12} \right]^{1/2}$

### 3.4. KỲ VỌNG TOÁN VÀ PHƯƠNG SAI CỦA MỘT SỐ LUẬT PHÂN PHỐI THƯỜNG GẶP

#### 3.4.1. Phân phối nhị thức

*a) Định nghĩa:* Trong một số ứng dụng kỹ thuật và yêu cầu giải quyết các vấn đề kinh tế, ta thường gặp các đại lượng ngẫu nhiên rời rạc  $X$  có xác suất tính theo công thức Bécnu-li, nghĩa là có bảng phân phối xác suất dạng:

$X$	0	1	2	...	$k$	...	$n$
$P$	$q^n$	$c_n^1 p q^{n-1}$	$c_n^2 p^2 q^{n-2}$	...	$c_n^k p^k q^{n-k}$	...	$p^n$

(với  $p+q = 1$ ).

Luật phân phối này được gọi là luật phân phối nhị thức. Rõ ràng, đối với luật phân phối nhị thức ta cũng có:

$$\sum_{k=0}^n p_k = \sum_{k=0}^n c_n^k p^k q^{n-k} = (p+q)^n = 1^n = 1$$

*b) Kỳ vọng toán:* Ta tính kỳ vọng toán của đại lượng ngẫu nhiên rời rạc  $X$  thuận theo luật phân phối nhị thức. Theo định nghĩa  $E(X)$  ta có:

$$E(X) = \sum_{k=0}^n x_k p_k = \sum_{k=0}^n k c_n^k p^k q^{n-k} \quad (37)$$

Lấy đạo hàm hai vế của nhị thức Newton  $(p+q)^n = \sum_{k=0}^n c_n^k p^k q^{n-k}$

theo  $p$  ta được:

$$\begin{aligned} \frac{d}{dp} [(p+q)^n] &= \frac{d}{dp} \left[ \sum_{k=0}^n c_n^k p^k q^{n-k} \right] \\ n(p+q)^{n-1} &= \sum_{k=0}^n k c_n^k p^{k-1} q^{n-k} \end{aligned} \quad (38)$$

Nhân cả hai vế của biểu thức (38) với  $p$ , ta có:

$$np(p+q)^{n-1} = \sum_{k=0}^n k c_n^k p^k q^{n-k} \quad (39)$$

Do  $(p+q)^{n-1} = 1^{n-1} = 1$  và so sánh (39) với (37) ta thu được biểu thức của kỳ vọng toán:

$$E(X) = np \quad (40)$$

**c) Phương sai:** Theo công thức tính phương sai của đại lượng ngẫu nhiên rời rạc  $X$  ta có:

$$D(X) = E(X^2) - [E(X)]^2 = E(X^2) - (np)^2 \quad (41)$$

ở đây:

$$E(X^2) = \sum_{k=0}^n x_k^2 p_k = \sum_{k=0}^n k^2 c_n^k p^k q^{n-k} \quad (42)$$

Để tính  $E(X^2)$ , ta lấy đạo hàm một lần nữa biểu thức (38) theo  $p$  cả hai vế:

$$\begin{aligned} \frac{d}{dp} [n(p+q)^{n-1}] &= \frac{d}{dp} \sum_{k=0}^n k c_n^k p^{k-1} q^{n-k} \\ n(n-1)(p+q)^{n-2} &= \sum_{k=0}^n k(k-1) c_n^k p^{k-2} q^{n-k} \end{aligned} \quad (43)$$



Nhân cả hai vế của (43) với  $p^2$ , ta được:

$$\begin{aligned} p^2 n(n-1) &= \sum_{k=0}^n k^2 c_n^k p^k q^{n-k} - \sum_{k=0}^n k c_n^k p^k q^{n-k} \\ &= \sum_{k=0}^n k^2 c_n^k p^k q^{n-k} - E(X) \end{aligned}$$

Suy ra:

$$E(X^2) = \sum_{k=0}^n k^2 c_n^k p^k q^{n-k} = p^2 n(n-1) + np \quad (44)$$

Thay (44) vào (41), ta thu được biểu thức của phương sai:

$$D(X) = p^2 n(n-1) + np - n^2 p^2 = np(1-p) = npq$$

Nghĩa là:  $D(X) = npq \quad (45)$

### 3.4.2. Phân phối Poátxông

**a) Định nghĩa:** Trong một số ứng dụng kỹ thuật ta thường gặp đại lượng ngẫu nhiên rời rạc  $X$  có bảng phân phối xác suất dạng:

$X$	0	1	2	...	$k$	...
$P$	$e^{-\lambda}$	$\frac{\lambda e^{-\lambda}}{1!}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$		$\frac{\lambda^k e^{-\lambda}}{k!}$	

(với  $\lambda$  hằng số).

Luật phân phối này được gọi là luật phân phối Poátxông.

Rõ ràng, đối với luật phân phối Poátxông ta cũng có:

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

**b) Kỳ vọng toán:** Theo định nghĩa kỳ vọng toán của đại lượng ngẫu nhiên rời rạc  $X$  ta có:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} x_k p_k = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda \end{aligned}$$

Nghĩa là:  $E(X) = \lambda$  (46)

**c) Phương sai:** Theo công thức tính phương sai của đại lượng ngẫu nhiên rời rạc  $X$  ta có:

$$D(X) = E(X^2) - [E(X)]^2 = E(X^2) - \lambda^2 \quad (47)$$

ở đây:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k^2 \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \sum_{k=1}^{\infty} (k+1) \cdot \frac{\lambda^{k+1} e^{-\lambda}}{(k+1)!} = \lambda \left[ \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \right] \\ &= [E(X) + \lambda] = \lambda(\lambda + 1) = \lambda^2 + \lambda \end{aligned} \quad (48)$$

Thay (48) vào (47), ta có:

$$D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Nghĩa là:

$$D(X) = \lambda \quad (49)$$

### 3.4.3. Phân phối chính qui (phân phối chuẩn)

**a) Định nghĩa:** Trong một số ứng dụng kỹ thuật ta thường gặp đại lượng ngẫu nhiên liên tục  $X$  có hàm mật độ phân phối xác suất  $f(x)$  dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (50)$$

(với  $\sigma$ ,  $a$  là các hằng số và  $\sigma > 0$ ).

Luật phân phối này được gọi là luật phân phối chính qui hay phân phối chuẩn.



Rõ ràng, đối với luật phân phối chính qui ta cũng có:

$$\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \sigma \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt \quad (51)$$

Tích phân Ôle:

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi} \quad (52)$$

Thay (52) vào (51), ta thu được:

$$\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \sigma \cdot \sqrt{2\pi} = 1$$

**b) Kỳ vọng toán:** Theo định nghĩa kỳ vọng toán của đại lượng ngẫu nhiên liên tục X, ta có:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t + a) e^{-\frac{t^2}{2}} \sigma dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sigma \left[ \int_{-\infty}^{+\infty} \sigma t e^{-\frac{t^2}{2}} dt + a \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt \right] \end{aligned} \quad (53)$$

Tích phân thứ nhất trong (53) là tích phân của hàm lẻ nên bằng 0, tích phân thứ hai trong (53) là tích phân Ôle nên (52) bằng  $\sqrt{2\pi}$ .

Vậy ta có:

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \sigma \left( 0 + a\sqrt{2\pi} \right) = a$$

Nghĩa là:

$$E(X) = a \quad (54)$$

Hằng số a trong (50) trùng với giá trị kỳ vọng toán E(X).

c) *Phương sai*: Theo công thức tính phương sai của đại lượng ngẫu nhiên liên tục X ta có:

$$D(X) = E(X^2) - [E(X)]^2 = E(X^2) - a^2 \quad (55)$$

ở đây:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t + a)^2 e^{-\frac{t^2}{2}} \sigma dt \\ &= \frac{1}{\sqrt{2\pi}} \left[ \int_{-\infty}^{+\infty} \sigma^2 t^2 e^{-\frac{t^2}{2}} dt + 2a\sigma \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt + \int_{-\infty}^{+\infty} a^2 e^{-\frac{t^2}{2}} dt \right] \end{aligned} \quad (56)$$

Tích phân thứ hai trong (56) là tích phân của hàm lẻ nên bằng 0, còn tích phân thứ ba trong (56) là tích phân Ore nên (52) bằng  $\sqrt{2\pi}$ .

Dẫn đến, ta có:

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 t^2 e^{-\frac{t^2}{2}} dt + a^2 \quad (57)$$

Xét tích phân:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 t^2 e^{-\frac{t^2}{2}} dt &= -\frac{2\sigma^2}{\sqrt{2\pi}} \left[ \frac{d}{d\varphi} \int_{-\infty}^{+\infty} e^{-\frac{\varphi t^2}{2}} dt \right]_{\varphi=1} \\ &= -\frac{2\sigma^2}{\sqrt{2\pi}} \left[ \frac{d}{d\varphi} \left( \varphi^{-\frac{1}{2}} \int_{-\infty}^{+\infty} e^{-\frac{\sqrt{\varphi} t^2}{2}} d(\sqrt{\varphi} t) \right) \right]_{\varphi=1} \\ &= -\frac{2\sigma^2}{\sqrt{2\pi}} \left[ \frac{d}{d\varphi} \left( \varphi^{-\frac{1}{2}} \cdot \sqrt{2\pi} \right) \right]_{\varphi=1} = -\frac{2\sigma^2}{\sqrt{2\pi}} \left( -\frac{1}{2} \sqrt{2\pi} \right) = \sigma^2 \end{aligned} \quad (58)$$



Thay (58) vào (57) ta có:

$$E(X^2) = \sigma^2 + a^2 \quad (59)$$

Thay (59) vào (55), ta thu được biểu thức của phương sai:

$$D(X) = \sigma^2 + a^2 - a^2 = \sigma^2$$

Nghĩa là:  $D(X) = \sigma^2$ .

Hằng số  $\sigma$  trong (50) trùng với giá trị của độ lệch quân phương  $\sigma(X) = \sqrt{D(X)}$ .

### 3.4.4. Phân phối mũ

*a) Định nghĩa:* Trong nhiều ứng dụng của lý thuyết xác suất, đặc biệt trong lý thuyết phục vụ đám đông, ta thường gặp đại lượng ngẫu nhiên liên tục  $X$  có hàm mật độ phân phối xác suất  $f(x)$  dạng:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{khi } x \geq 0 \\ 0 & \text{khi } x < 0 \end{cases} \quad (60)$$

(với  $\lambda$  là hằng số và  $\lambda > 0$ ).

Luật phân phối này được gọi là luật phân phối mũ.

Rõ ràng, đối với luật phân phối mũ ta cũng có:

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = - \int_0^{-\infty} e^{-t} dt = -(0 - 1) = 1$$

*b) Kỳ vọng toán:* Theo định nghĩa kỳ vọng toán của đại lượng ngẫu nhiên liên tục  $X$  ta có:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = - \int_0^{+\infty} x d(e^{-\lambda x}) \\ &= - x e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = 0 - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^{+\infty} = \frac{1}{\lambda} \end{aligned}$$

Nghĩa là:  $E(X) = \frac{1}{\lambda} \quad (61)$

c) *Phương sai*: Theo công thức tính phương sai của đại lượng ngẫu nhiên liên tục X, ta có:

$$D(X) = E(X^2) - [E(X)]^2 = E(X^2) - \frac{1}{\lambda^2} \quad (62)$$

ở đây:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = - \int_0^{+\infty} x^2 d(e^{-\lambda x}) \\ &= - x^2 e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx = \frac{2}{\lambda} E(X) = 2 \cdot \frac{1}{\lambda^2} \end{aligned} \quad (63)$$

Thay (63) vào (62) ta thu được:

$$D(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Nghĩa là: 
$$D(X) = \frac{1}{\lambda^2} \quad (64)$$

### 3.5. MỐT VÀ TRUNG VỊ

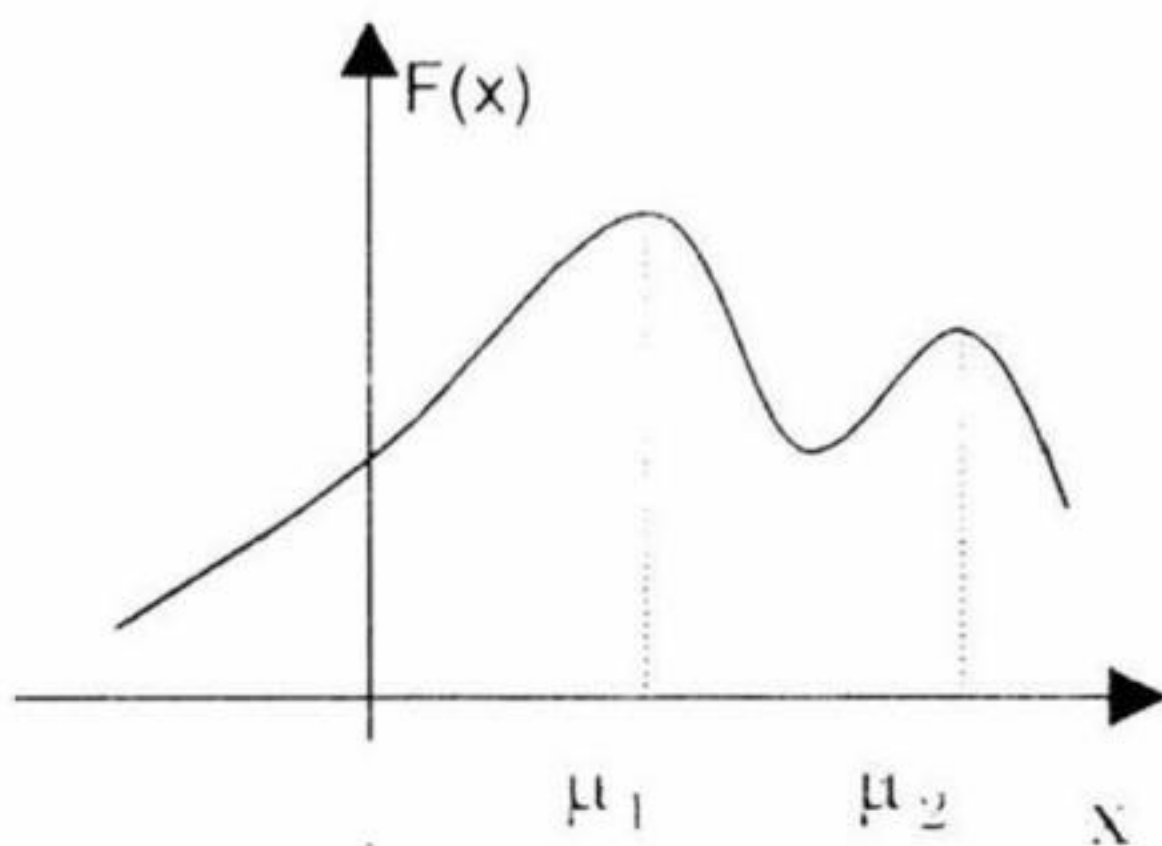
Ngoài kỳ vọng toán là đặc trưng quan trọng nhất về vị trí của đại lượng ngẫu nhiên, trên thực tế đôi khi người ta còn dùng đặc trưng khác về vị trí là một và trung vị.

**3.5.1. “Mốt”** của đại lượng ngẫu nhiên là trị số của đại lượng ngẫu nhiên có xác suất cực đại (đối với đại lượng rời rạc) hay là trị số có mật độ phân phối xác suất cực đại (đối với đại lượng liên tục). Ta ký hiệu mốt bởi chữ  $\mu$ .

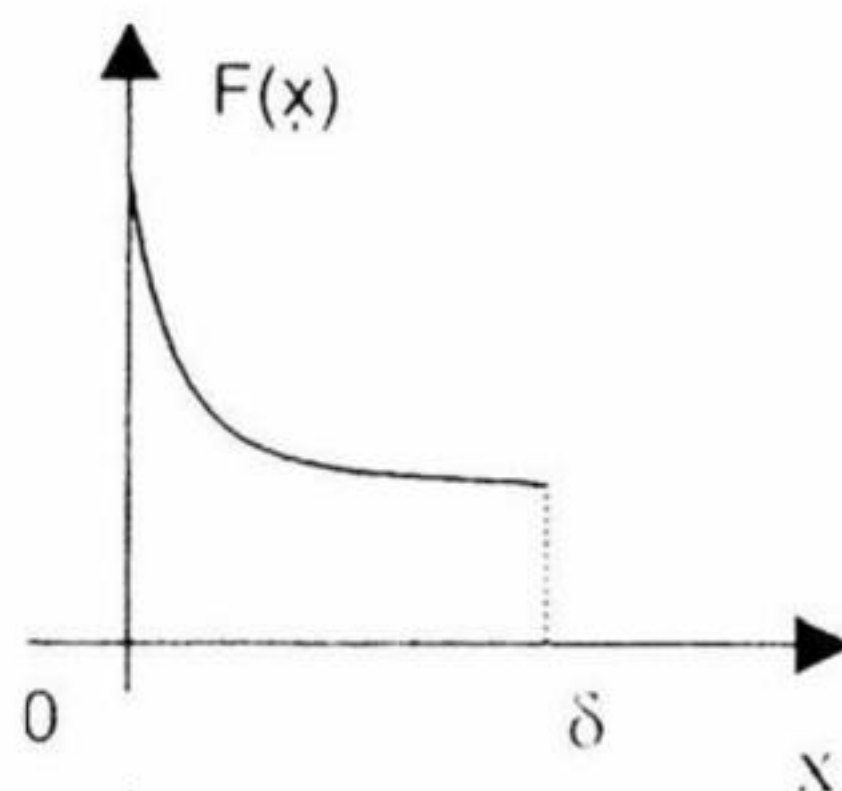
Nếu đường cong mật độ phân phối  $f(x)$  có nhiều điểm cực đại thì đại lượng ngẫu nhiên được gọi là có phân phối “nhiều mốt” (hình 13).

Nếu đường cong mật độ phân phối không có cực đại ở bên trong khoảng giá trị của đại lượng ngẫu nhiên thì ta có phân phối “không mốt” (hình 14).





Hình 13



Hình 14

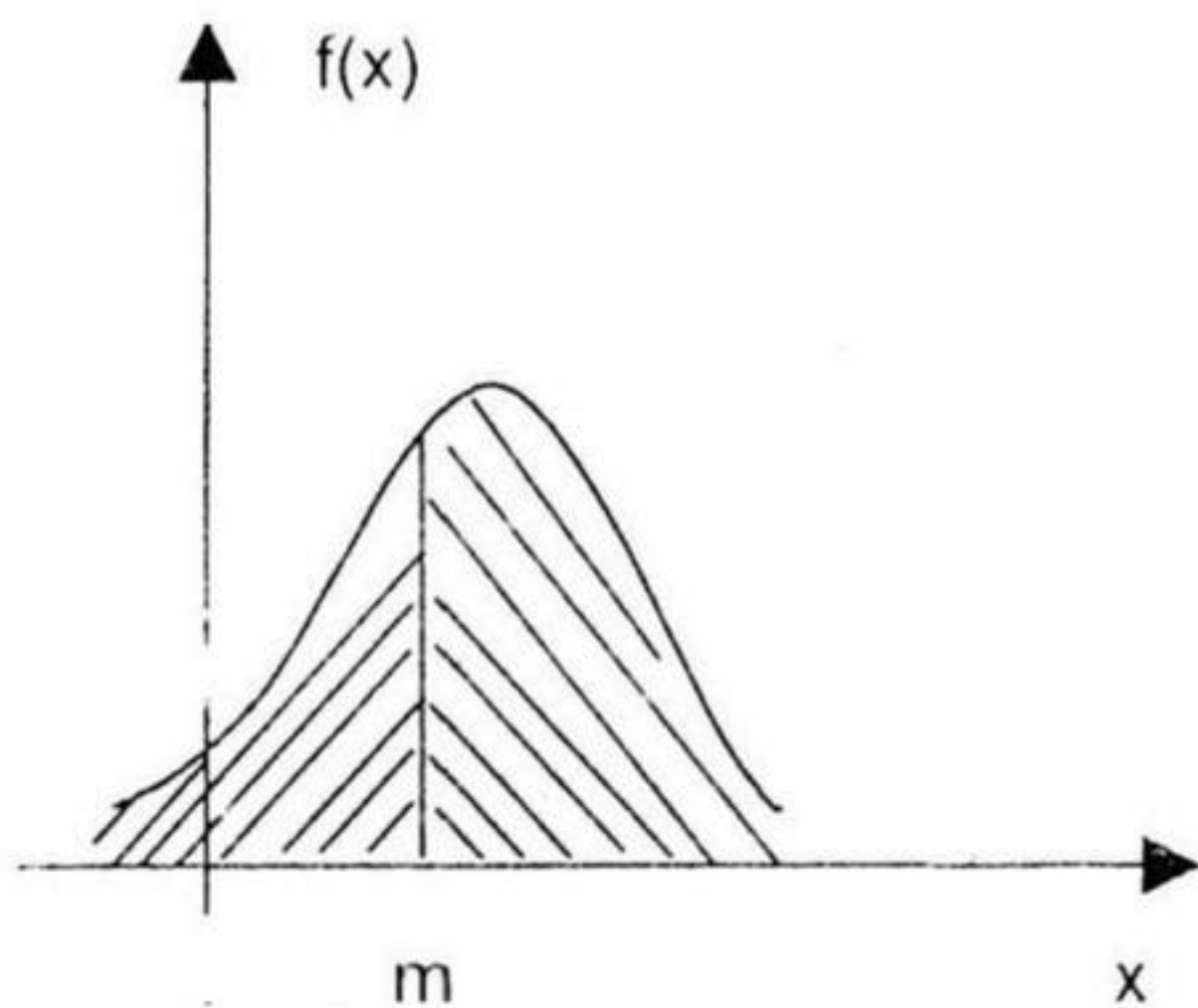
Nói chung một và kỳ vọng toán không trùng nhau. Hai đặc số này chỉ trùng nhau khi đường cong mật độ có cực đại và đối xứng.

**3.5.2.** Trung vị của đại lượng ngẫu nhiên  $X$  là trị số  $m$  của đại lượng ngẫu nhiên sao cho:

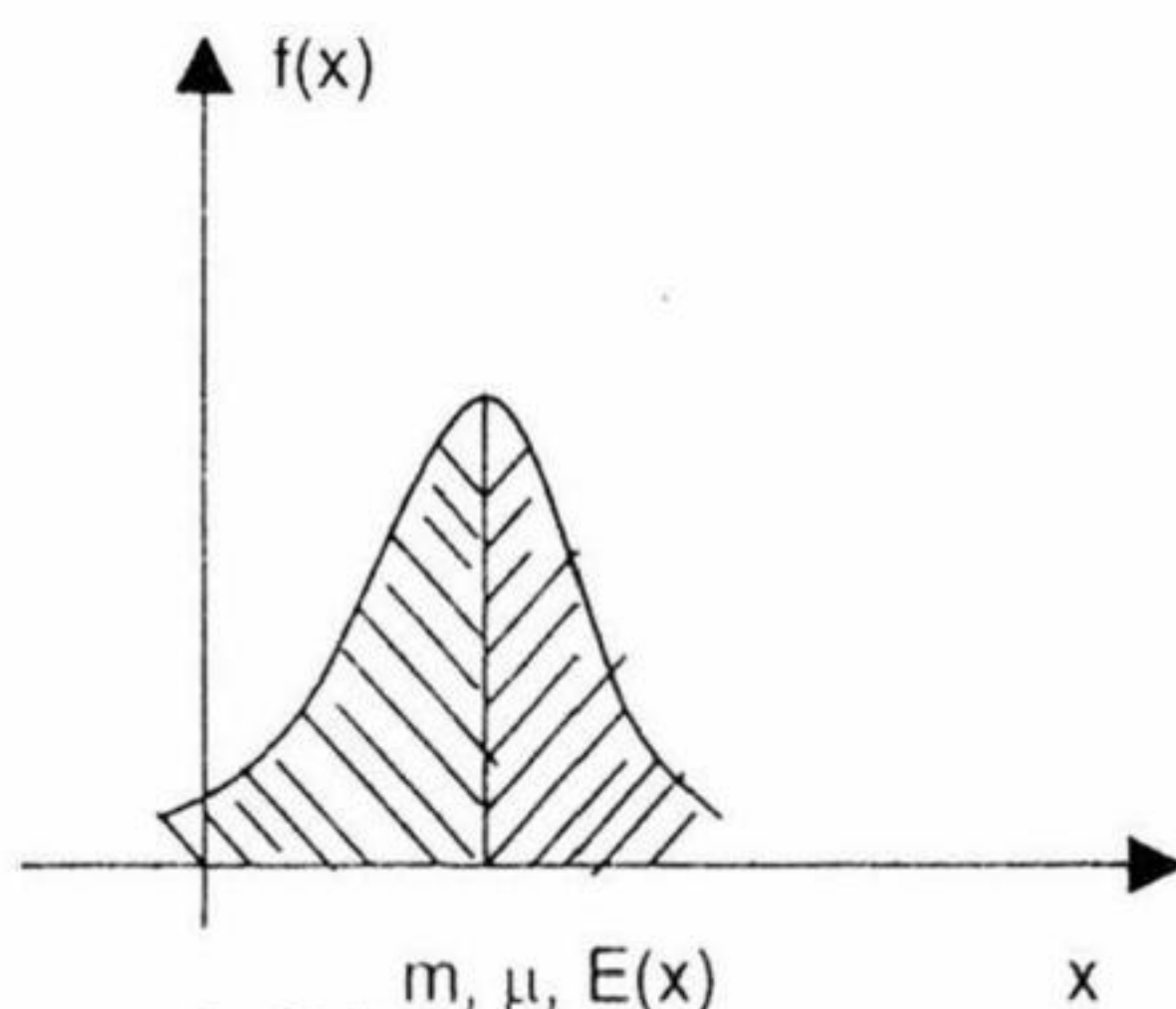
$$P(X < m) = P(X > m)$$

Về mặt hình học, trung vị là hoành độ của điểm tại đó diện tích giới hạn bởi đường cong mật độ phân phối được chia làm hai phần bằng nhau (hình 15).

Trong trường hợp phân phối đối xứng và có một thì ba đặc số: trung vị, một và kỳ vọng toán trùng nhau (hình 16).



Hình 15.



Hình 16

### 3.6. MÔMEN

Mômen trong lý thuyết xác suất là sự tổng quát hóa các khái niệm kỳ vọng toán và phương sai.

**Định nghĩa 1.** Mômen cấp  $k$  đối với điểm  $a$  của đại lượng ngẫu nhiên  $X$  là một số xác định như sau:

$$v_k(a) = E[(X - a)^k] \quad (65)$$

Nếu  $a = 0$  thì  $v_k(0)$  được gọi là mômen gốc. Để cho gọn ta ký hiệu mômen gốc là  $v_k$  thay cho  $v_k(0)$ .

Mômen gốc cấp 1 chính là kỳ vọng toán của biến ngẫu nhiên  $X$ :

$$v_1 = E(X^1) = E(X)$$

Nếu  $a = E(X)$  thì mômen cấp  $k$  ( $v_k(a)$ ) được gọi là *mômen trung tâm*; ký hiệu là  $\mu_k$ .

Dễ thấy rằng mômen trung tâm cấp 1 bằng 0, còn mômen trung tâm cấp 2 chính là phương sai:  $\mu_2 = D(X)$ .

Giữa các mômen gốc và mômen trung tâm có các liên hệ sau:

$$\mu_2 = D(X) = E(X^2) - E^2(X) = v_2 - v_1^2;$$

$$\begin{aligned} \mu_3 &= E[X - E(X)]^3 = E[X^3 - 3X^2E(X) + 3XE^2(X) - E^3(X)] \\ &= v_3 - 3v_1v_2 + 3v_1^3 - v_1^3 = v_3 - 3v_1v_2 - 2v_1^3 \end{aligned}$$

$$\mu_4 = v_4 - 4v_3v_1 + 6v_2v_1^2 - 3v_1^4$$

Trong số các mômen, thông dụng nhất là  $v_1$  (kỳ vọng toán),  $\mu_2$  (phương sai) rồi đến  $\mu_3$  và  $\mu_4$ . Sau đây ta sẽ nói về ý nghĩa của  $\mu_3$  và  $\mu_4$ .

Mômen trung tâm cấp ba  $\mu_3$  đặc trưng cho tính bất đối xứng của luật phân phối. Nếu như luật phân phối của đại lượng ngẫu nhiên đối xứng đối với kỳ vọng toán thì  $\mu_3$  bằng 0 (và tổng quát mọi mômen trung tâm lẻ đều bằng 0).

Thực vậy nếu đại lượng ngẫu nhiên  $X$  có phân phối xác suất đối xứng đối với kỳ vọng toán thì trong biểu thức:



$$\mu_3 = \sum_i (x_i - E(X))^3 p_i$$

mỗi số hạng dương sẽ bị triệt tiêu bởi một số hạng âm tương ứng có cùng trị số tuyệt đối, thành thử tổng bằng không. Cũng vậy tích phân:

$$\mu_3 = \int_{-\infty}^{\infty} [x - E(X)]^3 f(x) dx$$

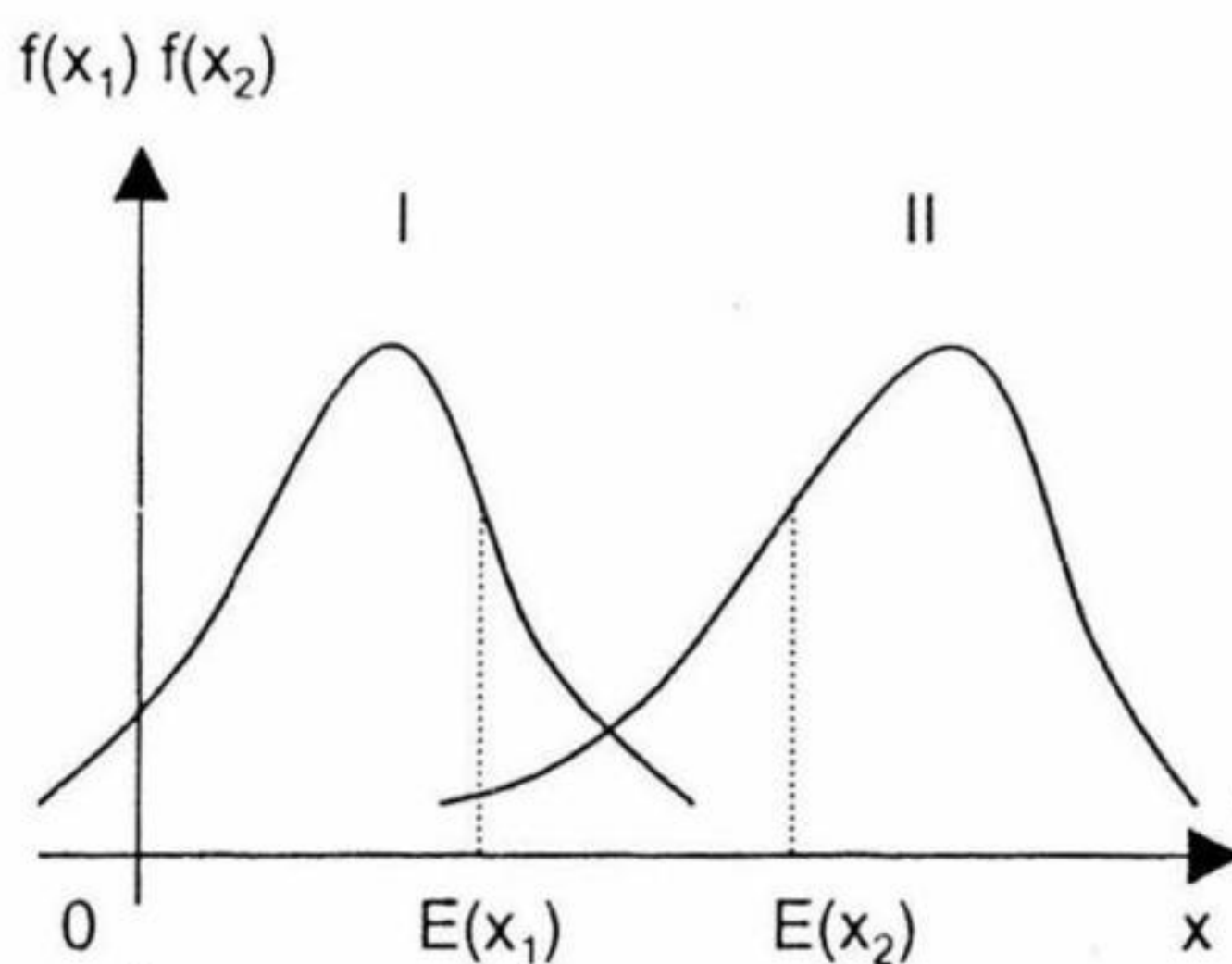
bằng 0 vì đó là tích phân của hàm số lẻ trong một khoảng đối xứng.

Từ đó ta suy ra rằng sự tồn tại  $\mu_3 \neq 0$  chứng tỏ tính bất đối xứng của luật phân phối. Cụ thể nếu  $\mu_3 > 0$  thì phân phối nặng về bên phải kỳ vọng toán, nếu  $\mu_3 < 0$  thì phân phối nặng về bên trái kỳ vọng toán.

Vì  $\mu_3$  có thứ nguyên bằng lũy thừa bậc ba thứ nguyên của  $X$  nên thường thay  $\mu_3$  bởi lượng:

$$S_3 = \frac{\mu_3}{\sigma^3}$$

được gọi là *hệ số bất đối xứng* (trong đó  $\sigma$  là độ lệch quân phương của  $X$ ).



**Hình 17**

Hình 17 vẽ hai luật phân phối bất đối xứng, một phân phối (đường cong II) có  $S_3 > 0$  và phân phối kia (đường cong I) có  $S_3 < 0$ .

Mômen trung tâm cấp bốn  $\mu_4$  đặc trưng cho độ nhọn của đường cong phân phối so với phân phối chính qui.

Đại lượng  $N_4 = \frac{\mu_4}{\sigma^4} - 3$  được gọi là “hệ số nhọn” của phân phối.

Đối với đại lượng ngẫu nhiên tuân theo phân phối chính qui  $N_4 = 0$ . Các đường cong nhọn hơn đường cong phân phối chính qui sẽ có  $N_4 > 0$  và tù hơn sẽ có  $N_4 < 0$ .

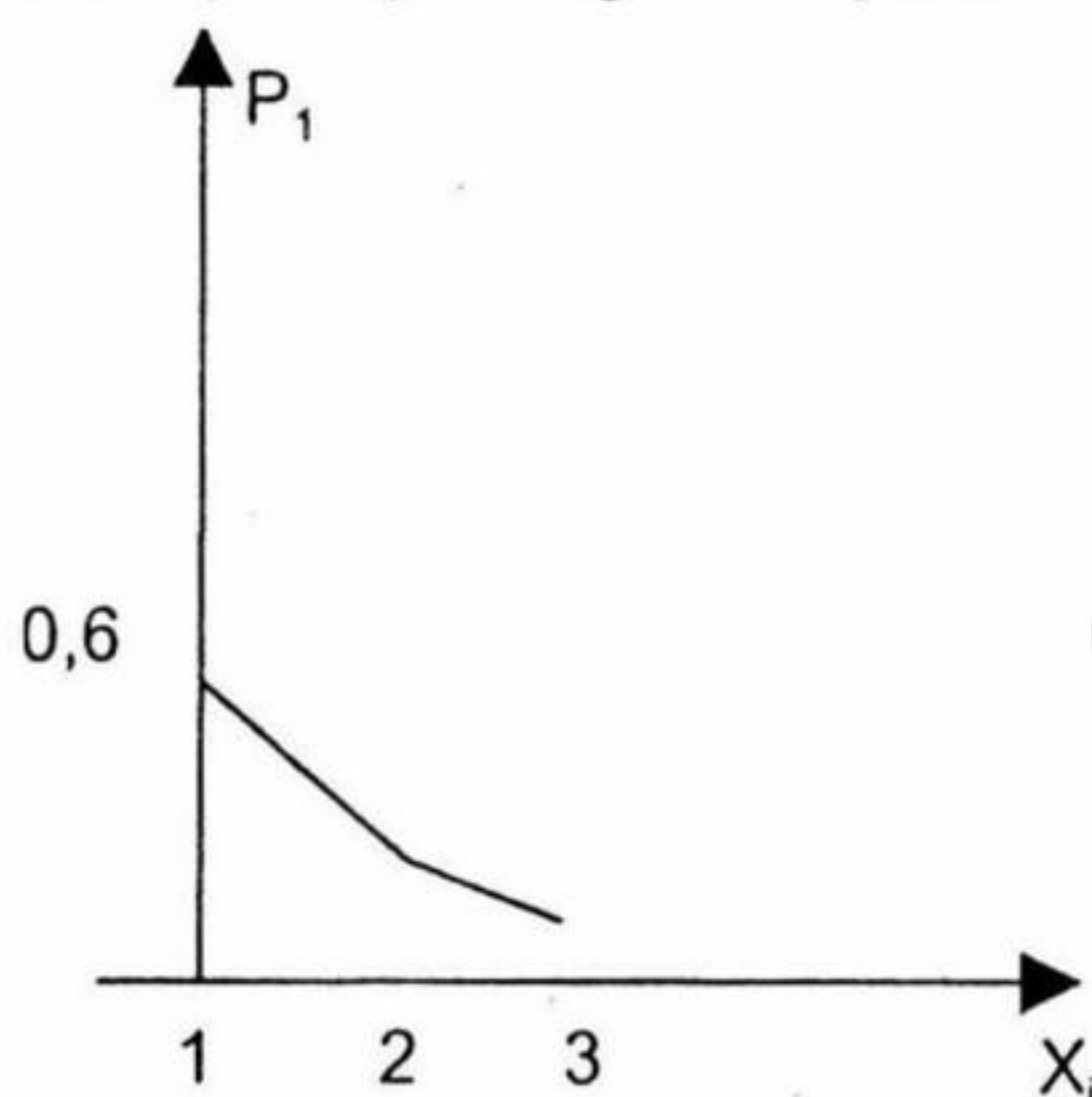
### BÀI TẬP CHƯƠNG 3

**Bài 1:** Một xạ thủ có 3 viên đạn được yêu cầu bắn lần lượt từng viên cho đến khi trúng thì dừng bắn. Tìm bảng phân phối xác suất của số đạn đã bắn, biết rằng xác suất bắn trúng của mỗi lần bắn là 0,6.

**Giải:** Gọi  $X$  là số đạn đã dùng, dễ thấy  $X$  có 3 giá trị là 1, 2 và 3. Nếu gọi  $A_i$  là sự kiện viên thứ  $i$  đã trúng ( $i=1, 2, 3$ ), ta có  $P(X=1)=P(A_1)$ ;  $P(X=2)=P(\bar{A}_1 A_2)$  và  $P(X=3)=P(\bar{A}_1 \bar{A}_2)$  lưu ý trong trường hợp  $X=3$  viên thứ 3 có thể trúng hoặc trượt). Từ đó luật phân phối xác suất của  $X$  cho bởi bảng:

$x_i$	1	2	3
$p_i$	0,6	0,24	0,16

Có thể biểu diễn luật này bằng đồ thị (xem hình 18).



Hình 18



**Bài 2:** Từ một lô hàng gồm 100 sản phẩm, trong đó có 10 phế phẩm, người ta chọn hủ hoá ra 5 sản phẩm để kiểm tra chất lượng. Lập bảng phân phối xác suất của số phế phẩm trong số sản phẩm chọn ra.

**Giải:** Lô hàng lấy ra có 5 sản phẩm nên số phế phẩm có giá trị nguyên từ 0 đến 5. Xác suất để đại lượng ngẫu nhiên  $X$  chỉ số phế phẩm trên nhận giá trị bằng  $i$ ,  $i = 0, 1, \dots, 5$  là  $P(X = i) = \frac{C_{10}^i C_{90}^{5-i}}{C_{100}^5}$ .

Với độ chính xác cỡ 0,001, ta có bảng sau:

$X_i$	0	1	2	3	4	5
$P_i$	0,583	0,340	0,070	0,007	0,000	0,000

Cần chú ý rằng  $P(X=4)=P(X=5) = 0$  không có nghĩa là không thể xảy ra biến cố có 4 hoặc 5 phế phẩm trong lô hàng.

Ở đây các biến cố ấy rất khó xảy ra trên thực tế do xác suất của chúng quá bé.

**Bài 3:** Tiến hành các xét nghiệm độc lập cho đến khi có kết quả dương tính. Gọi  $X$  là số xét nghiệm đã tiến hành và biết rằng xác suất dương tính của mỗi xét nghiệm là 0,5. Tìm:

- Bảng phân phối xác suất của  $X$ ;
- Đồ thị phân phối;
- Số xét nghiệm có xác suất lớn nhất.

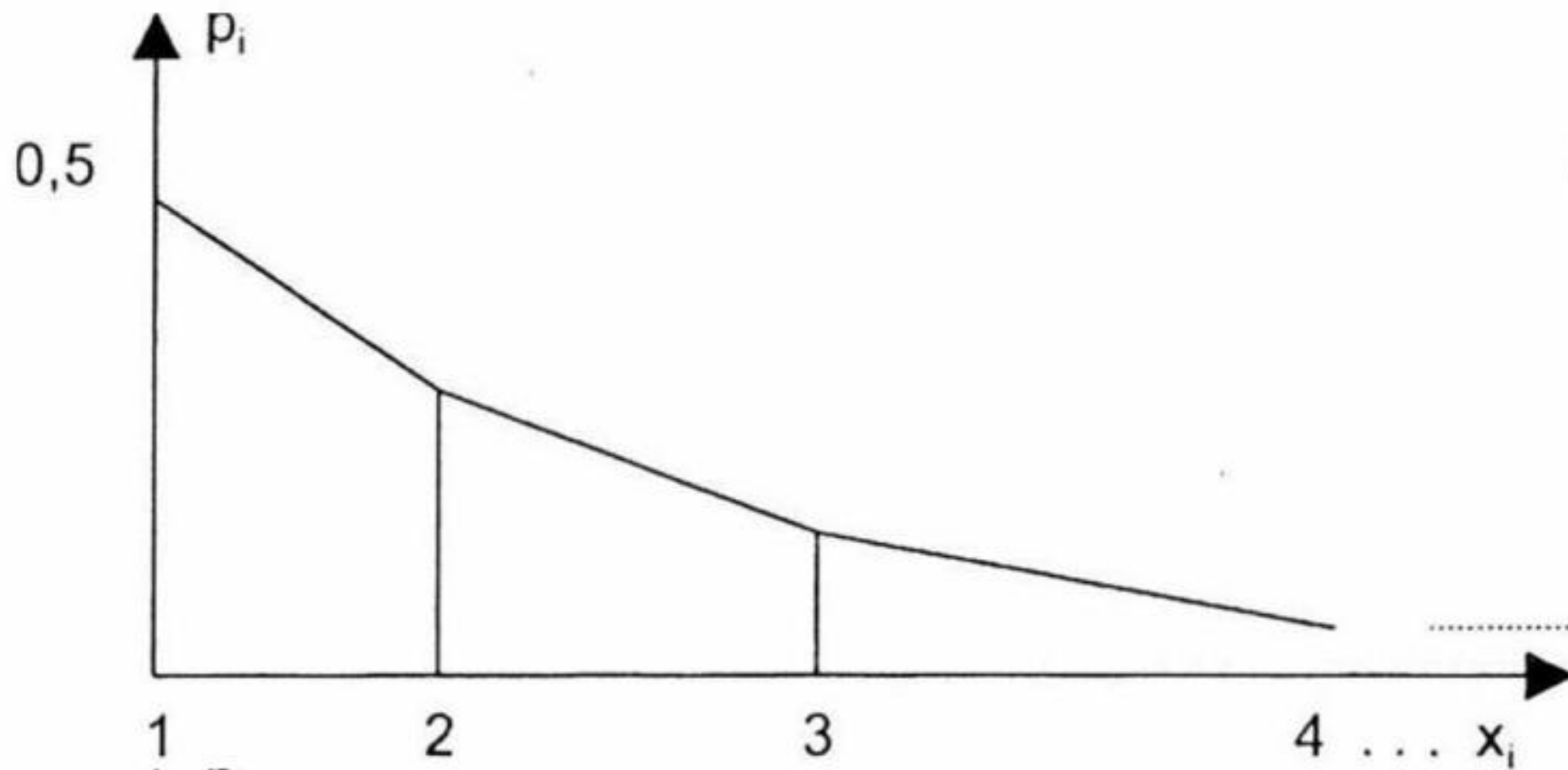
**Giải:**

a) Giá trị của  $X$  sẽ là 1, 2, 3 . . . . Dễ thấy  $P(X=i)=q^{x-1}p$ , với  $p = 0,5$  và  $q = 1-p = 0,5$ . Từ đó:

$X_i$	1	2	3	...	$n$	...
$P_i$	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{1}{2^3}$	...	$\frac{1}{2^n}$	...

b) Đồ thị phân phối trên hình 19.

c) Số xét nghiệm có khả năng xuất hiện lớn nhất (số chắc nhất) là 1.



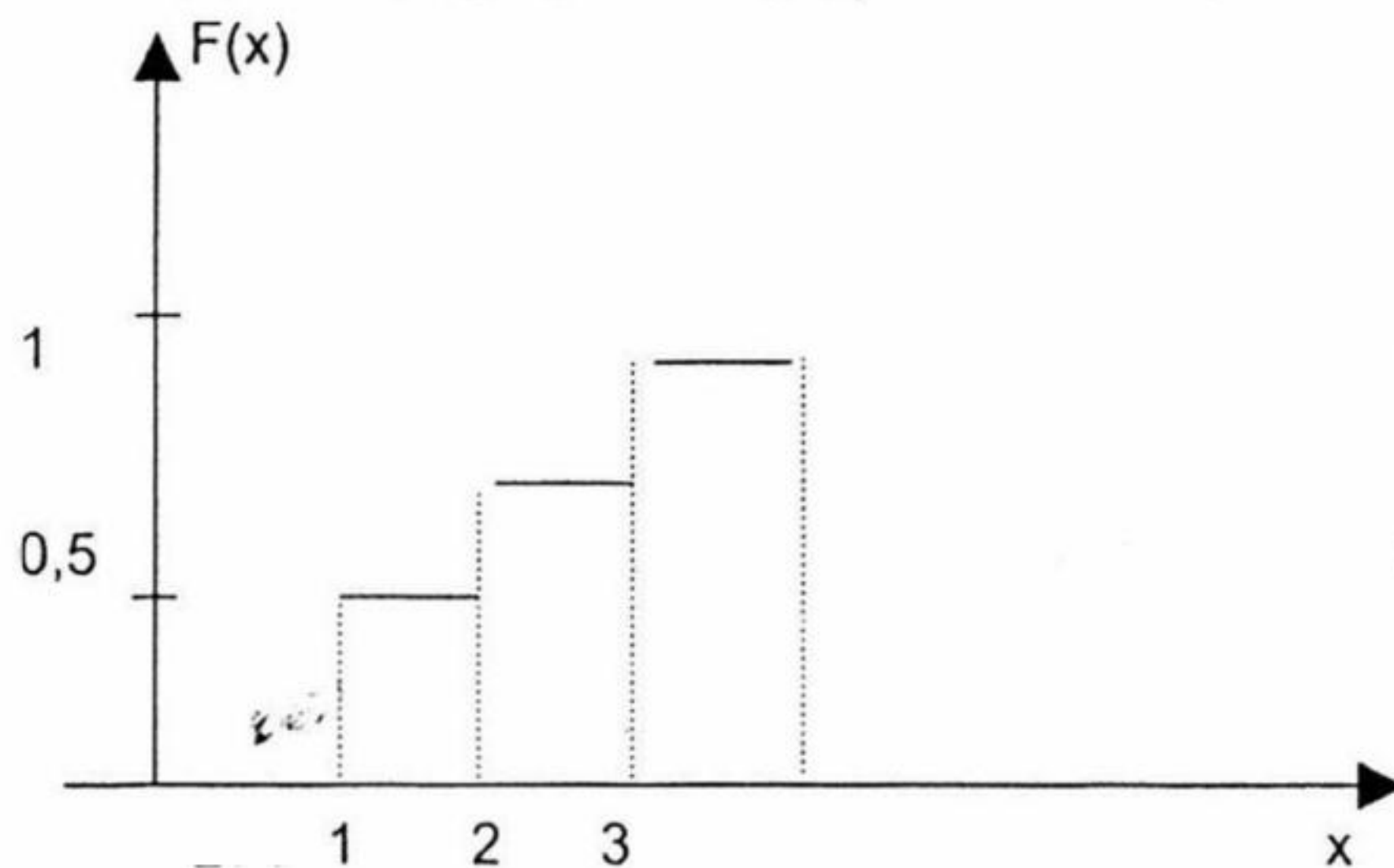
**Hình 19**

**Bài 4:** Xác định hàm phân phối xác suất của đại lượng ngẫu nhiên  $X$  trong bài 1.

**Giải:** Theo định nghĩa hàm phân phối ta có ngay:

$$F(x) = \begin{cases} 0, & x \leq 1, \\ 0,6, & 1 < x \leq 2, \\ 0,84, & 2 < x \leq 3, \\ 1, & x > 3. \end{cases}$$

Đồ thị của hàm  $F(x)$  có dạng bậc thang (xem hình 20).



**Hình 20**



**Bài 5:** Hàm phân phối của một đại lượng ngẫu nhiên liên tục  $X$  có dạng:

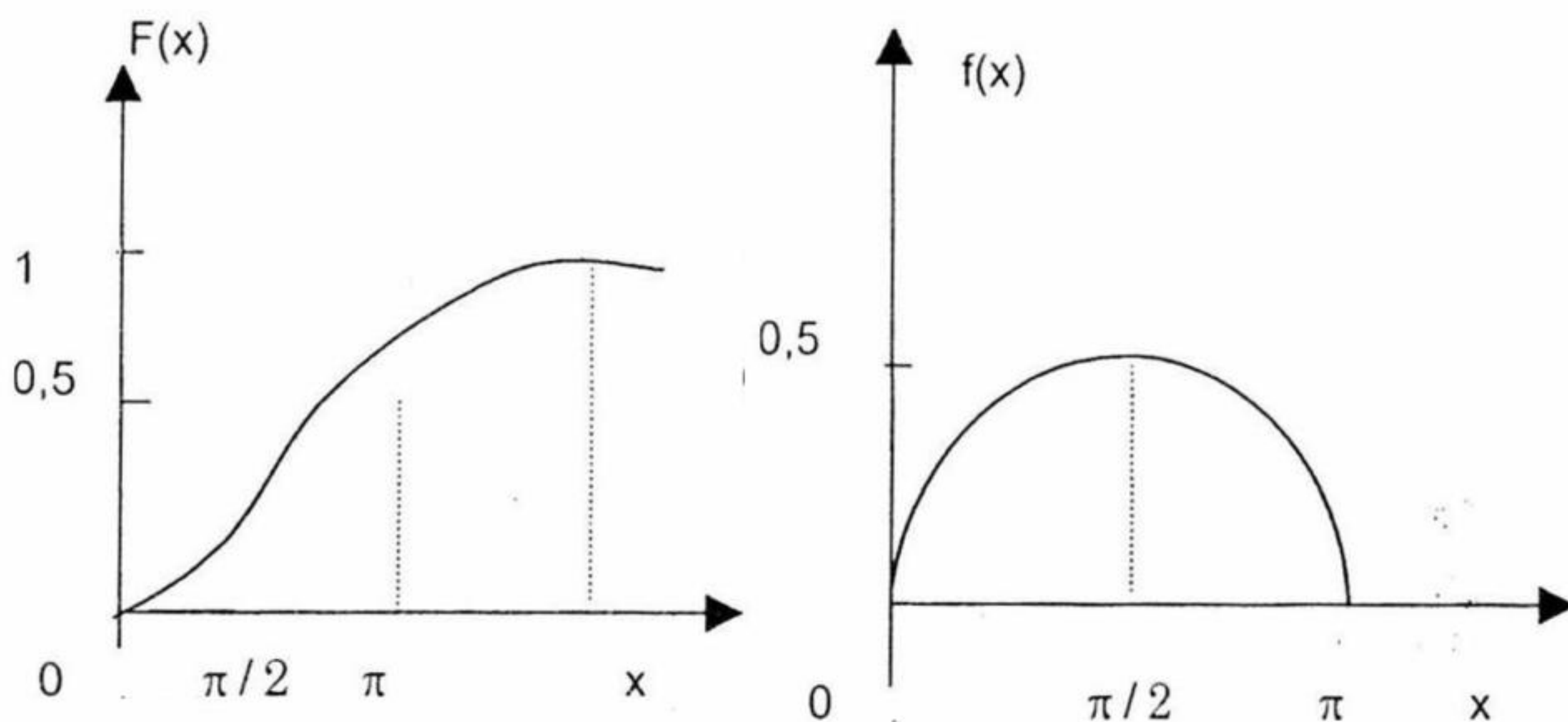
$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2} - \frac{1}{2} \cos x, & 0 < x \leq \pi, \\ 1, & x > \pi. \end{cases}$$

Hãy xác định hàm mật độ của  $X$  và vẽ đồ thị của  $F(x)$  và  $f(x)$ .

**Giải:** Theo định nghĩa hàm mật độ ta có:

$$f(x) = \begin{cases} 0, & x \leq 0, \\ \frac{1}{2} \sin x, & 0 < x \leq \pi, \\ 0, & x > \pi. \end{cases}$$

Đồ thị  $F(x)$  và  $f(x)$  vẽ trên hình 21.



**Hình 21**

**Bài 6:** Cho hàm phân phối của đại lượng ngẫu nhiên  $X$  có dạng:

$$F(x) = a + b \arctg x \quad (-\infty < x < +\infty)$$

Tìm  $a$ ,  $b$  và hàm mật độ  $f(x)$ .

**Giải:** Dùng tính chất 3 của hàm phân phối ta có:

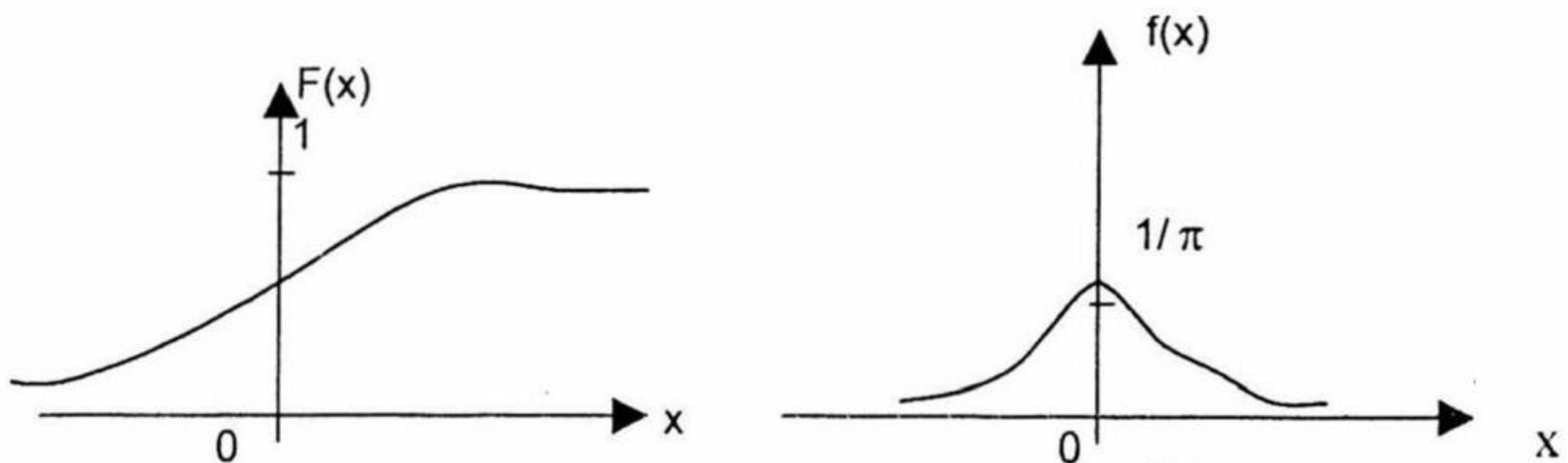
$$\begin{cases} \lim_{x \rightarrow +\infty} (a + b \arctg x) = 0 \\ \lim_{x \rightarrow +\infty} (a + b \arctg x) = 1 \end{cases} \quad \text{hay} \quad \begin{cases} a - \frac{b\pi}{2} = 0 \\ a + \frac{b\pi}{2} = 1 \end{cases}$$

suy ra:  $a = \frac{1}{2}$ ;  $b = \frac{1}{\pi}$  và  $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctg x$ .

Ta có hàm mật độ:

$$f(x) = F'(x) = \frac{1}{\pi(1+x^2)}.$$

Có thể xem đồ thị của các hàm này trên hình 22:



**Hình 22**

**Bài 7:** Đại lượng ngẫu nhiên  $X$  có hàm mật độ sau đây:

$$f(x) = \begin{cases} 0, & x < 1 \\ \frac{A}{x^2}, & x \geq 1 \end{cases}$$

Hãy xác định:

- Hệ số  $A$ ;
- Hàm phân phối  $F(x)$ ;
- $P(2 < X < 3)$  - xác suất  $X$  rơi vào khoảng  $(2, 3)$ .
- Xác suất để trong 4 phép thử độc lập đại lượng  $X$  đều không lấy giá trị trên khoảng  $(2, 3)$ .



**Giải:**

a) Việc xác định A dựa vào tính chất  $\int_{-\infty}^{+\infty} f(x) dx = 1$ . Ta có :

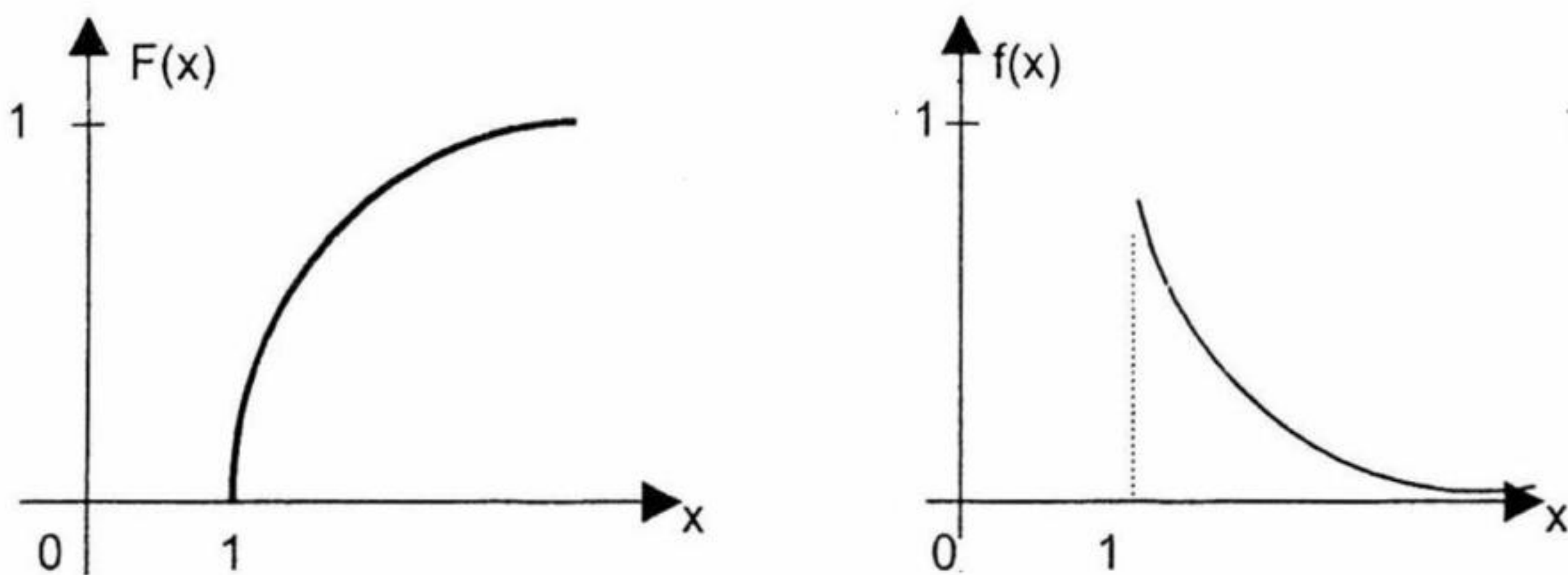
$$\int_{-\infty}^{+\infty} f(x) dx = - \lim_{a \rightarrow +\infty} \left( \frac{A}{x} \Big|_1^a \right) = A$$

Từ đó  $A = 1$ .

b) Từ tính chất của hàm mật độ:

$$F(x) = \int_{-\infty}^x f(x') dx' = \begin{cases} 0, & x < 1 \\ \frac{x-1}{x}, & x \geq 1 \end{cases}$$

Đồ thị của các hàm trên vẽ trong hình 23.



**Hình 23**

$$c) P(2 < X < 3) = F(3) - F(2) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

d) Xác suất để X không rơi vào khoảng (2,3) (không lấy giá trị trên khoảng đó) trong một phép thử bằng  $1 - \frac{1}{6} = \frac{5}{6}$ , vậy trong bốn phép

$$\text{thử bằng } \left( \frac{5}{6} \right)^4 \approx 0,48$$

**Bài 8:** Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm, trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000 đôla, còn tiền đóng là 10 đôla. Hỏi lợi nhuận của công ty đó là bao nhiêu?

**Giải:** Rõ ràng lợi nhuận là biến ngẫu nhiên  $X$  với giá trị là 10 đôla (nếu người bảo hiểm không chết) và  $-990$  đôla (nếu người đó chết). Bảng phân phối xác suất tương ứng.

$X$	+10	-990
$P(X)$	0,992	0,008

Từ đó  $E(X) = 10 \cdot 0,992 + (-990) \cdot 0,008 = 2$ . Ta thấy lợi nhuận trung bình là một số dương và công ty bảo hiểm có thể làm ăn có lợi.

**Bài 9:** Đại lượng ngẫu nhiên  $X$  có bảng phân phối xác suất.

$x_i$	-1	0	1
$p_i$	0,2	0,3	0,5

Hãy tính  $E(X)$ ,  $E(X^2)$ ,  $D(X)$ .

**Giải:** Theo các công thức (26), (29), (30).

$$E(X) = -1 \cdot 0,2 + 0 \cdot 0,3 + 1 \cdot 0,5 = 0,3.$$

$$E(X^2) = (-1)^2 \cdot 0,2 + 0^2 \cdot 0,3 + 1^2 \cdot 0,5 = 0,7 \neq [E(X)]^2.$$

$$D(X) = E(X^2) - [E(X)]^2 = 0,7 - 0,09 = 0,61.$$

Đề ý rằng có thể tính  $D(X)$  theo công thức (27) nhưng việc tính sẽ khó khăn hơn nhiều.

**Bài 10.** Cho đại lượng ngẫu nhiên  $X$  có hàm mật độ

$$f(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{a^2} e^{-x^2/a^2}, & x \geq 0. \end{cases}$$

Tìm Med và Mod của  $X$  ( $a > 0$ ) (tức tìm trung vị và mốt của  $X$ ).



**Giải:** Đặt  $m = \text{Med}X$ , ta phải có:

$$F(m) = \frac{1}{2}$$

Với  $F(x)$  là hàm phân phối của  $X$  ta có  $F(x) = \int_{-\infty}^x f(t)dt$  hay:

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-x^2/2a^2}, & x \geq 0. \end{cases}$$

$$\text{Từ đó } 1 - e^{-m^2/2a^2} = \frac{1}{2}, \text{ suy ra } m = \sqrt{2 \ln 2} \cdot a.$$

Nếu đặt  $m = \text{Mod}X$ , ta phải có  $f'(m) = 0$ , tức là:

$$f'(m) = \left( \frac{1}{a^2} - \frac{m^2}{a^4} \right) e^{-m^2/2a^2} = 0.$$

Từ đó suy ra  $m = a$ .

**Bài 11:** Xác suất trúng đích của một khẩu súng là  $p$ . Tiến hành bắn liên tiếp trong điều kiện không đổi cho đến khi có  $k$  phát súng trúng đích thì thôi bắn. Tìm kỳ vọng toán của số lần bắn cần thiết.

**Giải:** Gọi  $X$  là số lần bắn cần thiết thì  $X$  nhận giá trị  $x$  nguyên nào đó nếu trước đó đã có  $k-1$  lần bắn trúng và lần thứ  $x$  đó cũng bắn trúng. Để ý số  $k-1$  lần có thể tùy ý (không phân biệt thứ tự trong số  $x-1$  lần). Vì lần cuối cùng (thứ  $x$ ) phải trúng nên:

$$P(X = x) = C_{x-1}^{k-1} p^{k-1} q^{x-k} p = C_{x-1}^{k-1} p^k q^{x-k}.$$

Để ý rằng nếu  $x < k$  thì  $P(X=x) = 0$  (theo điều kiện đầu bài). Từ đó theo công thức (19).

$$E(X) = \sum_{x=k}^{\infty} x C_{x-1}^{k-1} p^k q^{x-k} = kp^k \sum_{x=k}^{\infty} C_{x-1}^{k-1} q^{x-k}$$

Ta đổi biến  $x-k = i$  hay  $x = k+i$  và:

$$E(X) = kp^k \sum_{i=0}^{\infty} C_{i+k-1}^{k-1} q^i = \frac{kp^k}{(1-q)^k} = \frac{k}{p}$$

**Bài 12:** Cho hàm mật độ của một đại lượng ngẫu nhiên  $X$  như sau:

$$f(x) = \begin{cases} 0, & x > 2 \\ ax^2, & 0 \leq x < 1 \\ a(2-x)^2, & 1 \leq x < 2 \\ 0, & x \geq 2. \end{cases}$$

Xác định hệ số  $a$ , sau đó tính các mômen gốc và mômen trung tâm tới cấp 4.

**Giải:** Theo tính chất của kỳ vọng toán để tìm hệ số  $a$  ta phải giải:

$$a \int_0^1 x^2 dx + a \int_1^2 (2-x)^2 dx = 1$$

Từ đó  $a = \frac{3}{2}$ . Đầu tiên ta tìm các mômen gốc:

$$v_1 = \frac{3}{2} \int_0^1 x^3 dx + \frac{3}{2} \int_1^2 x(2-x)^2 dx = 1$$

$$v_2 = \frac{3}{2} \int_0^1 x^4 dx + \frac{3}{2} \int_1^2 x^2(2-x)^2 dx = 1,1$$

$$v_3 = \frac{3}{2} \int_0^1 x^5 dx + \frac{3}{2} \int_1^2 x^3(2-x)^2 dx = 1,3$$

$$v_4 = \frac{3}{2} \int_0^1 x^6 dx + \frac{3}{2} \int_1^2 x^4(2-x)^2 dx = 1 \frac{22}{35}$$

Các mômen trung tâm:

$$\mu_1 = 0; \mu_2 = v_2 - v_1^2 = 1,1 - 1 = 0,1$$

$$\mu_3 = v_3 - 3v_1v_2 + 2v_1^3 = 0$$

$$\mu_4 = v_4 - 4v_1v_3 + 6v_1^2v_2 - 3v_1^4 = \frac{1}{35}$$

**Bài 13:** Tìm kỳ vọng toán, phương sai và độ lệch quân phương của đại lượng ngẫu nhiên  $X$  cho bởi bảng phân phối:



X	1	3	5
P	0,6	0,3	0,1

**Giải:** Theo định nghĩa  $E(X)$  ta có:

$$E(X) = \sum_{i=1}^n x_i p_i = 1.1,6 + 3.0,3 + 5.0,1 = 2$$

Theo công thức tính  $D(X)$ , ta có:

$$D(X) = E(X^2) - [E(X)]^2 = 1^2.0,6 + 3^2.0,3 + 5^2.0,1 - 2^2 = 5,8 - 4 = 1,8$$

Theo định nghĩa độ lệch quân phương, ta có:

$$\sigma \equiv \sigma(X) = \sqrt{D(X)} = \sqrt{1,8} \approx 1,34$$

**Bài 14:** Tìm kỳ vọng toán của số lần xuất hiện biến cố A trong một phép thử, nếu xác suất xuất hiện biến cố A trong phép thử bằng p.

**Giải:** Gọi X là đại lượng ngẫu nhiên chỉ số lần xuất hiện biến cố A trong một phép thử. Ta có bảng phân phối xác suất.

X	0	1
P	(1-p)	p

Theo định nghĩa  $E(X)$ , ta có:

$$E(X) = 0.(1-p) + 1.p = p$$

**Bài 15:** Mật độ phân phối của đại lượng ngẫu nhiên liên tục  $f(x)$  có dạng:

$$f(x) = Ax^2 \cdot e^{-Kx}, \quad (\text{với } K > 0, \quad 0 \leq x \leq +\infty)$$

Hãy tính hệ số A và tính hàm phân phối xác suất  $F(X)$ .

**Giải:** Để tìm A ta xuất phát từ tính chất của hàm mật độ phân phối:

$$1 = \int_{-\infty}^{+\infty} f(x) dx = A \int_0^{+\infty} x^2 e^{-Kx} dx$$

Tích phân từng phần ta có:

$$\begin{aligned}
1 &= A \left[ -x^2 \frac{e^{-Kx}}{K} \Big|_0^\infty + \frac{1}{K} \int_0^\infty 2xe^{-Kx} dx \right] \\
&= \frac{2A}{K} \int_0^\infty xe^{-Kx} dx = \frac{2A}{K} \left[ -X \frac{e^{-Kx}}{K} \Big|_0^\infty + \frac{1}{K} \int_0^\infty e^{-Kx} dx \right] \\
&= \frac{2A}{K^2} \left( -\frac{1}{K} \right) e^{-Kx} \Big|_0^\infty = \frac{2A}{K^3}
\end{aligned}$$

Suy ra:  $A = \frac{K^3}{2}$ .

Để tính  $F(x)$ , ta xuất phát từ công thức:

$$F(x) = \int_{-\infty}^x f(x') dx'$$

Nếu  $x < 0$ , ta có  $f(x)=0$ , suy ra  $F(x) = 0$ .

Nếu  $0 \leq x \leq +\infty$ , ta có:

$$F(x) = \int_0^x Ax'^2 e^{-Kx'} dx'$$

Tích tích phân trên ta được:

$$F(x) = 1 - \frac{e^{-Kx}}{2} (K^2 x^2 + 2Kx + 2)$$



## Chương 4

# MẪU THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

### 4.1. KHÔNG GIAN MẪU

Để nghiên cứu một hay nhiều tính chất nào đấy của một tập hợp nhiều vật thể ít khi người ta có thể mang tất cả các vật thể ra để nghiên cứu vì số lượng lớn và có khi thí nghiệm làm hư hỏng vật thể. Vì vậy người ta tìm cách lấy ra một số trong tất cả các vật thể nói trên để nghiên cứu rồi từ đó kết luận về các tính chất cần thiết của tất cả các vật thể ban đầu.

Để nghiên cứu chiều dài của hạt lúa thuộc một giống nào đó thì người ta không thể mang tất cả các hạt lúa ra đo được; để biết thời gian có thể làm việc của bóng điện không thể mang tất cả bóng điện đã sản xuất ra để thí nghiệm được.

Tập hợp tất cả các vật thể ban đầu được gọi là tập hợp chính, hay còn gọi là tập hợp toàn bộ. Tập hợp các vật thể lấy ra được gọi là mẫu. Số phần tử của mẫu được gọi là số lượng của mẫu (còn gọi là cỡ mẫu).

Bằng một phương pháp có thể lấy ra nhiều mẫu khác nhau có cùng một số lượng. Tập hợp tất cả các mẫu có thể lấy ra được gọi là không gian mẫu và mỗi mẫu được coi là một điểm của không gian mẫu.

Người ta phân biệt hai loại mẫu là mẫu có lập và mẫu không lập. Xét tập hợp chính gồm  $N$  vật thể  $a_1, a_2, \dots, a_N$  và mẫu gồm  $n$  vật thể ( $n \leq N$ ) ký hiệu là  $a_{j1}, a_{j2}, \dots, a_{jn}$ . Cách lấy mẫu thứ nhất như sau: trước tiên ta lấy hù hoạ (ngẫu nhiên) một phần tử trong tập hợp chính và gọi là  $a_{j1}$ , sau đó bỏ phần tử lấy được trở lại tập hợp chính, rồi lại lấy



hút một phần tử làm phần tử thứ hai  $a_{j2}$ , lại bỏ phần tử thứ 2 trở lại tập hợp chính rồi lấy hút một phần tử làm phần tử thứ ba  $a_{j3}$ , và cứ tiếp tục như thế đến phần tử thứ  $n$  thì được  $a_{jn}$ . Như vậy một phần tử có thể được lấy nhiều lần. Mẫu được lấy ra như vậy được gọi là có lặp. Cách lấy mẫu thứ hai như sau: lấy từ tập hợp chính phần tử thứ nhất và không bỏ nó trở lại nữa, tiếp tục lấy phần tử thứ hai và không bỏ nó trở lại, tiếp tục lấy phần tử thứ ba v: v. . . Như thế sẽ không có một phần tử nào được chọn 2 lần. Mẫu này được gọi là mẫu không lặp.

Nếu số phần tử trong tập hợp chính là  $N$  và mẫu có số lượng  $n$  thì có  $N^n$  cách lấy mẫu khác nhau trong trường hợp mẫu có lặp. Thật vậy vì mỗi phần tử trong mẫu có lặp có thể được chọn  $N$  cách, vậy  $n$  phần tử có  $N^n$  cách chọn khác nhau.

Tương tự ta có  $N(N-1) \dots (N-n+1)$  cách lấy mẫu trong trường hợp không lặp. Trong cách tính này ta có chú ý đến thứ tự các phần tử trong mẫu. Nếu ta không để ý đến thứ tự các phần tử trong mẫu (nghĩa là hai mẫu được coi như nhau nếu chúng cùng chứa những phần tử như nhau) thì số mẫu không lặp có thể lấy ra là  $\frac{A_N^n}{n!}$  với

$$A_N^n = N(N-1) \dots (N-n+1).$$

Nếu  $N$  lớn và  $n$  nhỏ thì  $\frac{A_N^n}{N^n}$  khá gần 1. Điều đó chứng tỏ khi  $N$  lớn và số lượng của mẫu bé thì việc lấy mẫu có lặp cũng có kết quả gần với việc lấy mẫu không lặp.

Muốn cho từ mẫu lấy được có thể suy ra tương đối chính xác tính chất của tập hợp chính thì mẫu phải tiêu biểu. Mẫu được coi là tiêu biểu nếu người ta lấy mẫu một cách ngẫu nhiên, nghĩa là lấy thế nào để mọi phần tử của tập hợp chính có thể rơi vào mẫu với xác suất như nhau.

Với các định lý của lý thuyết xác suất trong các chương sau, từ tính chất của vật thể ở mẫu ta thấy rằng có thể suy ra tính chất của các vật thể trong tập hợp chính với độ chính xác cho trước.



Trong thực tế có nhiều cách lấy mẫu. Sau đây ta sẽ trình bày phương pháp lấy mẫu ngẫu nhiên đơn giản là phương pháp thường dùng.

Giả sử tập hợp chính có  $N$  phần tử và cần lấy mẫu với lượng  $n$  ( $n \leq N$ ).

Người ta đánh số tất cả các phần tử thuộc tập hợp chính từ 1 đến  $N$ , đồng thời làm  $N$  chiếc thẻ đánh số từ 1 đến  $N$ . Từ  $N$  thẻ rút hù họa một chiếc, ghi lấy số của nó rồi lấy phần tử trong tập hợp chính có số trùng với số vừa lấy được làm phần tử đầu tiên của mẫu. Bỏ thẻ vừa rút ra được trở lại tập hợp các thẻ, sau đó rút hù họa lần thứ 2, đọc số của thẻ vừa rút ra và lấy phần tử có mang số thứ tự này làm phần tử thứ 2 của mẫu. Lại bỏ thẻ vào và rút ra lần thứ 3. Cứ làm như vậy cho đến lần thứ  $n$  thì lấy được đủ  $n$  phần tử của mẫu, và như vậy ta có một mẫu có lập. Nếu thẻ đã rút ra không được trở lại nữa thì ta có mẫu không lập.

Nếu số phần tử  $N$  của tập hợp chính lớn thì không thể dùng các thẻ để lấy mẫu được mà người ta thường dùng bảng số ngẫu nhiên.

Sau đây là ví dụ dùng bảng số ngẫu nhiên của Kadiróp.

Mỗi số thuộc bảng này gồm 4 chữ số (4 phần tử), 5 số lập thành một nhóm, 10 nhóm lập thành một cột và mỗi trang có 10 cột. Như vậy mỗi trang có  $5 \cdot 10 \cdot 10 = 500$  số ngẫu nhiên. Bảng số gồm nhiều trang.

Trong mỗi số, các phần tử (4 phần tử) được chọn một cách ngẫu nhiên trong các số từ 0 đến 9 (Bảng V phần phụ lục).

Bây giờ giả sử một tập hợp chính có 543 phần tử và cần lấy mẫu có số lượng là 12.

Trước hết ta đánh số tất cả các phần tử ở tập hợp chính từ 0 đến 542. Sau đó dùng bảng số ngẫu nhiên chọn hù họa một trang, trong trang ấy lấy hù họa một cột rồi lại chọn hù họa một số trong cột vừa chọn. Ba chữ số bất kỳ của số vừa chọn được lấy làm số của phần tử thứ nhất của mẫu (nếu ba chữ số đó lập thành một số bé hơn hoặc bằng 542. Ta còn phải lấy 11 phần tử nữa, có thể theo một cách đọc số



bất kỳ đã định trước. Ví dụ ta có thể đọc từ trên xuống dưới hoặc đọc ngược từ dưới lên trên theo cột, hoặc đọc ngang theo hàng từ trái sang phải hoặc từ phải sang trái. Từ đó có kết quả cần tìm chẳng hạn số đầu tiên đọc được là 2157 ở trang đầu cột thứ 3, hàng 2 từ trên xuống, ta bỏ một chữ số bất kỳ trong 4 chữ số trên đi, chẳng hạn ta bỏ số 7 và được số 215 làm phân tử đầu tiên của mẫu. Ta đọc từ trên xuống dưới, đọc theo cột ứng với vị trí các chữ số đã được chọn, bỏ các số lớn hơn 542 đi, ta được các số tiếp theo: 250; 062; 381; 164; 084; 438; 050; 486; 501; 364; 031. Và như vậy được mẫu cần thiết.

Nếu lấy mẫu không lặp thì đối với các số trùng nhau ta chỉ giữ lại một số. Sau đây để đơn giản khi ta nói đến mẫu thì có nghĩa là mẫu có lặp và lấy theo phương pháp ngẫu nhiên đơn giản.

Nếu lấy mẫu từ tập hợp chính để nghiên cứu đại lượng ngẫu nhiên  $X$  và được kết quả là  $(X_1, X_2, \dots, X_n)$  thì ta còn nói là đã lấy mẫu  $(X_1, X_2, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$ .

## 4.2. PHÂN PHỐI MẪU VÀ PHÂN PHỐI CHÍNH XÁC

Giả sử mẫu  $(X_1, X_2, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$  có hàm phân phối  $F(x)$  (còn gọi là phân phối chính xác của  $X$ ). Ở đây vì chưa biết  $F(x)$  nên ta căn cứ vào mẫu để tìm một hàm số nào đấy gần với  $F(x)$ .

Ta lập một hàm phân phối xác suất mới  $F_n(x)$  bằng cách đặt:

$$F_n(x) = \frac{n_x}{n} \quad (1)$$

trong đó  $n_x$  chỉ số phân tử của mẫu có trị số nhỏ hơn  $x$  ( $X_i < x$ ),  $F_n(x)$  được gọi là hàm phân phối mẫu của đại lượng ngẫu nhiên  $X$ . Rõ ràng là sau khi đã lấy mẫu rồi thì phân phối này được xác định hoàn toàn. Theo (1),  $F_n(x)$  là tần suất của sự kiện ( $X < x$ ) ứng với  $n$  phép thử độc lập.

Ta thấy (1) tương đương với:



$$P(X = X_i) = \frac{1}{n} \quad (i=1, 2, \dots, n) \quad (2)$$

Rõ ràng hàm phân phối mẫu cũng là một hàm phân phối xác suất. Từ phân phối mẫu  $F_n(x)$ , theo các định nghĩa ta có:

Kỳ vọng mẫu của X:

$$E_n(X) = \sum_{i=1}^n \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (3)$$

Phương sai mẫu của X:

$$D_n(X) = \sum_{i=1}^n \frac{1}{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4)$$

Ở đây  $E_n(X)$  và  $D_n(X)$  chỉ kỳ vọng và phương sai của X được tính theo phân phối mẫu  $F_n(x)$ , khác với  $E(X)$  và  $D(X)$  được tính theo  $F(x)$ . Để đơn giản, sau này ta dùng ký hiệu  $EX$  thay cho  $E(X)$  và  $DX$  thay cho  $D(X)$ .

Ta cũng tính được các mômen mẫu  $v_k$ , các mômen trung tâm mẫu như sau:

$$\begin{aligned} v_k &= \frac{1}{n} \sum_{i=1}^n X_i^k \\ \mu_k &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \end{aligned} \quad (5)$$

và các hệ số bất đối xứng, độ nhọn mẫu:

$$\begin{aligned} S_3 &= \frac{\mu_3}{(\mu_2)^{3/2}} \\ N_1 &= \frac{\mu_4}{\mu_2^2} - 3 \end{aligned} \quad (6)$$

Cần phân biệt hàm phân phối mẫu  $F_n(x)$  của đại lượng ngẫu nhiên X, với hàm phân phối chính xác  $F(x)$  của đại lượng ngẫu nhiên X.

Như ta đã biết  $F_n(x)$  bằng tần suất của sự kiện  $(X < x)$  còn  $F(x) = P(X < x)$  bằng xác suất của sự kiện đó.

Do đó, với  $n$  khá lớn người ta dùng  $F_n(x)$  thay cho  $F(x)$  khi  $F(x)$  chưa biết.

Để cho đơn giản, từ nay về sau chúng ta nói hàm phân phối của  $X$  là nói hàm phân phối chính xác của  $X$ , còn khi cần đến  $F_n(x)$  ta sẽ nói rõ là phân phối mẫu (hay phân phối thực nghiệm) của  $X$ .

### 4.3. PHÂN PHỐI XÁC SUẤT CỦA ĐẠI LƯỢNG THỐNG KÊ TRÊN KHÔNG GIAN MẪU

#### 4.3.1. Phân phối xác suất của đại lượng thống kê

Giả sử ta có mẫu  $(X_1, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$  có mật độ  $f(x)$ . Rõ ràng  $X_i$  ( $i=1, \dots, n$ ) có thể nhận mọi trị số trên miền giá trị của  $X$  với luật phân phối  $f(x_i)$  và vì mẫu có lặp nên người ta qui ước các  $X_i$  ( $i=1, \dots, n$ ) cùng có phân phối xác suất, và chúng độc lập với nhau. Nếu  $X$  là đại lượng ngẫu nhiên (một chiều) thì  $(X_1, \dots, X_n)$  là một vectơ ngẫu nhiên (đại lượng ngẫu nhiên  $n$  chiều),  $(X_1, \dots, X_n)$  có thể nhận các trị  $(x_1, \dots, x_n)$  trong đó  $x_i$  ( $i=1, \dots, n$ ) là các hằng số ứng với một mẫu cụ thể đã được lấy ra,  $(X_1, \dots, X_n)$  có mật độ là  $f(x_1) \times \dots \times f(x_n)$ . Một hàm số  $g(X_1, \dots, X_n)$  bất kỳ với biến là  $(X_1, \dots, X_n)$  được gọi là đại lượng thống kê trên không gian mẫu. Vì  $(X_1, \dots, X_n)$  là đại lượng ngẫu nhiên (nhiều chiều) nên  $g(X_1, \dots, X_n)$  cũng là đại lượng ngẫu nhiên.

Ví dụ, kỳ vọng mẫu:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

và phương sai mẫu:

$$\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

cũng như các đặc trưng mẫu nói ở 4.2 đều là các đại lượng thống kê trên không gian mẫu.



Cho  $Y=g(X_1, \dots, X_n)$  là đại lượng thống kê và  $f(x_1) \times \dots \times f(x_n)$  là mật độ của  $(X_1, \dots, X_n)$ . Vấn đề đặt ra là hãy tìm hàm phân phối  $H(y)$  của  $Y$ .

Ta có thể chứng minh được:

$$H(y) = \int_G f(x_1) \dots f(x_n) dx_1 \dots dx_n$$

với:

$$G = \{(x_1, \dots, x_n) : g(x_1, \dots, x_n) < y\}$$

Trên đây ta mới nêu một phương pháp để từ luật phân phối xác suất của  $(X_1, \dots, X_n)$  suy ra luật phân phối của  $g(X_1, \dots, X_n)$ . Tất nhiên còn nhiều phương pháp nữa như dùng hàm đặc trưng, đổi biến số v.v... để giải quyết tùy theo bài toán cụ thể (nghĩa là tùy dạng của  $f(x)$  và của  $g(X_1, \dots, X_n)$ ).

#### 4.3.2. Phân phối xác suất của một số đại lượng thống kê thường gặp

a) Ta biết nếu  $X$  và  $Y$  là hai đại lượng ngẫu nhiên độc lập với nhau và có phân phối chuẩn  $N(a, \sigma_1^2)$  và  $N(b, \sigma_2^2)$  thì  $X+Y$  có phân phối chuẩn  $N(a+b, \sigma_1^2 + \sigma_2^2)$  (Để cho tiện từ đây ta dùng chữ phân phối chuẩn cũng có nghĩa như phân phối chính quy và dùng ký hiệu  $N(a, \sigma^2)$  để chỉ phân phối chuẩn có kỳ vọng  $a$  và phương sai  $\sigma^2$ ). Từ nhận xét trên suy ra :

**Định lý 1.** Nếu mẫu  $(X_1, \dots, X_n)$  được lấy từ đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\theta, \sigma^2)$  thì  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  có phân phối chuẩn

$$N\left(\theta, \frac{\sigma^2}{n}\right).$$

b) **Phân phối  $\chi^2$**  (đọc là khi bình phương).

**Định nghĩa.** Nếu  $X_i$  ( $i=1, \dots, n$ ) là các đại lượng ngẫu nhiên độc lập và cùng có phân phối chuẩn  $N(0,1)$ , thì phân phối xác suất của

$U = \sum_{i=1}^n X_i^2$  được gọi là phân phối  $\chi^2$  với  $n$  bậc tự do ( $n$  là một số nguyên  $\geq 1$ ) và ký hiệu là  $\chi_n^2$ .

**Định lý 2.** Hàm mật độ của đại lượng ngẫu nhiên có phân phối  $\chi^2$  với  $n$  bậc tự do có dạng:

$$f(u) = \frac{1}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)} u^{n/2-1} e^{-u/2} \quad (6)$$

với  $u > 0$  và  $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx \quad (a > 1)$ .

*Chứng minh:* Ta chỉ cần chứng minh hàm đặc trưng của  $U$  trùng với hàm đặc trưng của đại lượng ngẫu nhiên có mật độ (2). Hàm đặc trưng của  $X_j^2$  là :

$$\varphi(\lambda) = E e^{i\lambda X_j^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\lambda x^2} \cdot e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2 (1-2i\lambda)} dx$$

đặt  $z = x\sqrt{1-2i\lambda}$  .

$$\varphi(\lambda) = \frac{1}{(1-2i\lambda)^{1/2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz = \frac{1}{(1-2i\lambda)^{1/2}}$$

vì  $\int_{-\infty}^{+\infty} e^{-z^2/2} dz = \sqrt{2\pi}$  . Các đại lượng  $X_i$  ( $i=1, 2, \dots, n$ ) độc lập và cùng

phân phối xác suất nên  $U$  có hàm đặc trưng là:

$$\psi(\lambda) = \frac{1}{(1-2i\lambda)^{n/2}}$$

Bây giờ ta tìm hàm đặc trưng của đại lượng ngẫu nhiên có mật độ (2).

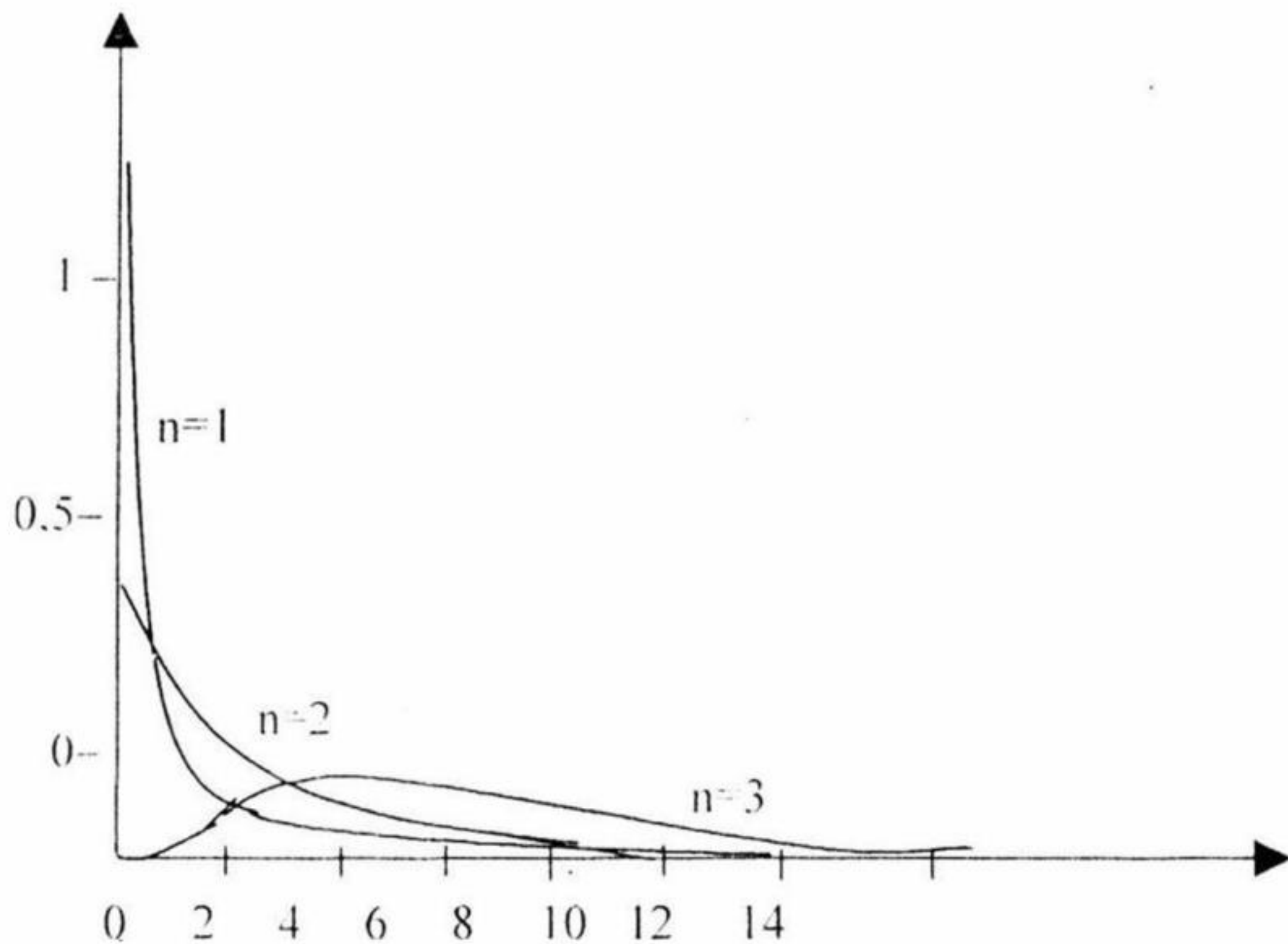


$$\begin{aligned}
 g(\lambda) &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} u^{\frac{n}{2}-1} e^{-\frac{u}{2}} e^{i\lambda u} du \\
 &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} u^{\frac{n}{2}-1} e^{-\frac{u}{2}(1-2i\lambda)} du
 \end{aligned}$$

đặt  $z = \frac{u}{2}(1-2i\lambda)$ , ta có:

$$\begin{aligned}
 g(\lambda) &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} \left(\frac{2z}{1-2i\lambda}\right)^{\frac{n}{2}-1} e^{-z} \frac{2}{1-2i\lambda} dz = \\
 &= \frac{1}{(1-2i\lambda)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} z^{\frac{n}{2}-1} e^{-z} dz = \frac{1}{(1-2i\lambda)^{\frac{n}{2}}}
 \end{aligned}$$

Như vậy  $\psi(\lambda) = g(\lambda)$ , đó là điều cần chứng minh.



Hình 24

Ở trong hình 24, ta có đồ thị của mật độ phân phối  $\chi^2$  với bậc tự do  $n=1;2;3$ . Bây giờ xét mẫu  $(X_1, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$ .

**Định lý 3.** Nếu  $X$  có phân phối chuẩn  $N(a, \sigma^2)$ , thì  $\frac{n-1}{\sigma^2} S^2$  có phân phối  $\chi^2$  với  $(n-1)$  bậc tự do.

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 ;$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Để chứng minh định lý 3 người ta chỉ ra rằng  $\frac{n-1}{\sigma^2} S^2$  có thể biểu diễn dưới dạng tổng của  $(n-1)$  đại lượng ngẫu nhiên độc lập với nhau, có cùng phân phối chuẩn  $N(0,1)$ , tuy nhiên việc làm này tương đối dài nên ta không đưa vào đây.

### c) Phân phối Stiuđon

#### Định nghĩa:

Đại lượng ngẫu nhiên  $t$  có dạng:

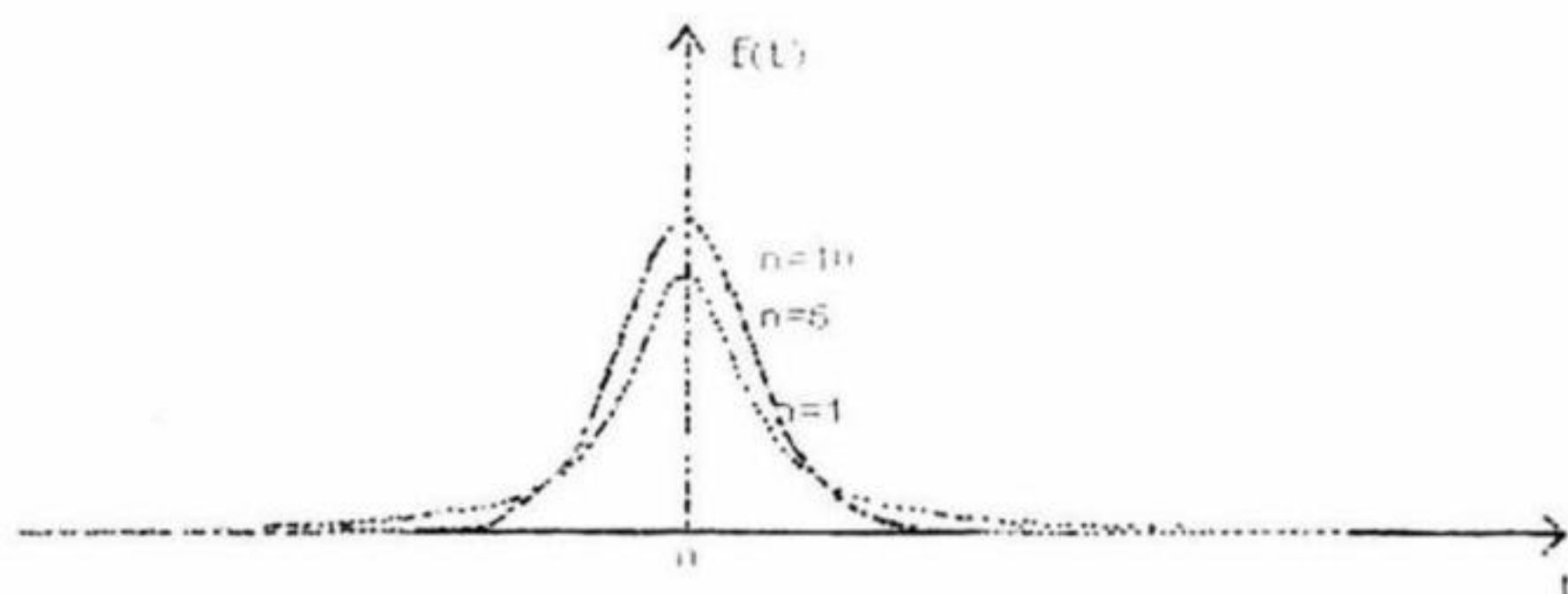
$$t = \frac{z}{\sqrt{u}} \sqrt{n} \quad (7)$$

Với  $z$  có phân phối chuẩn  $N(0,1)$ ,  $u$  có phân phối  $\chi^2$  với  $n$  bậc tự do,  $u$  và  $z$  độc lập với nhau, thì phân phối của  $t$  được gọi là phân phối Stiuđon với  $n$  bậc tự do. Mật độ của phân phối Stiuđon với  $n$  bậc tự do là:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \cdot \sqrt{\pi \cdot n}} \cdot \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}} \quad (8)$$



Trong hình 25 có mật độ của phân phối Stiuđon với bậc tự do  $n=1, 5, 10$ .



**Hình 25**

Xét mẫu  $(X_1, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$ .

**Định lý 4.** Nếu  $X$  có phân phối chuẩn  $N(a, \sigma^2)$ , thì  $t = \frac{\bar{X} - a}{S} \sqrt{n}$  ( $n$

là số lượng của mẫu,  $\bar{X}$  - trung bình mẫu,  $S = \sqrt{S^2}$ ,

$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ) có phân phối Stiuđon với  $(n-1)$  bậc tự do.

#### **d) Phân phối $F$ của Phisơ-Snêđêcô**

**Định nghĩa:**

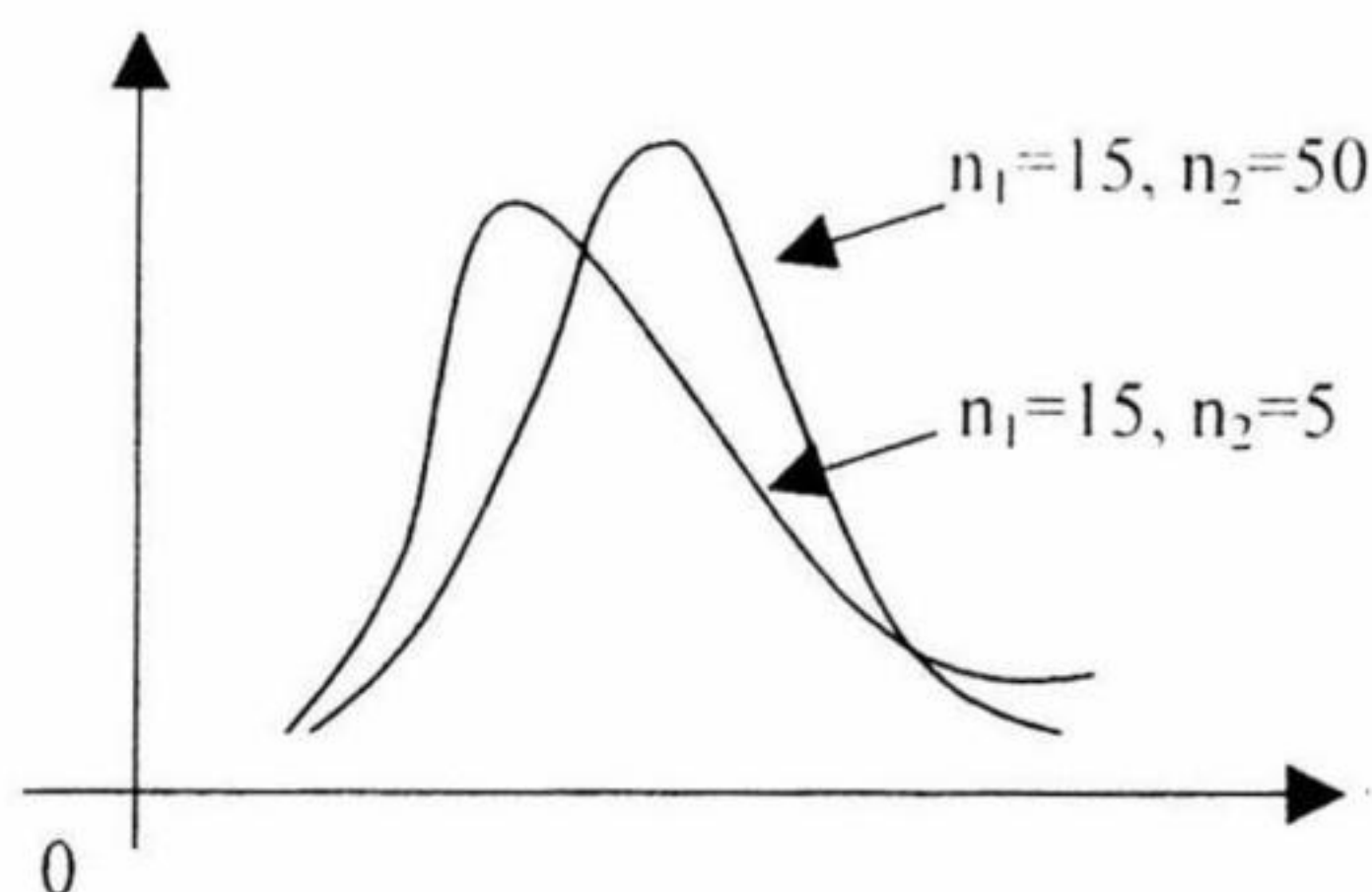
Nếu đại lượng ngẫu nhiên  $F$  có dạng:

$$F = \frac{\frac{U_1^2}{n_1}}{\frac{U_2^2}{n_2}} \quad (9)$$

trong đó  $U_1^2$  và  $U_2^2$  là các đại lượng ngẫu nhiên độc lập với nhau có phân phối  $\chi^2$  với  $n_1$  và  $n_2$  bậc tự do thì phân phối của  $F$  được gọi là phân phối  $F$  với  $n_1, n_2$  bậc tự do ( ký hiệu  $F_{n_1, n_2}$  ). Mật độ của phân phối  $F$  với  $n_1, n_2$  bậc tự do là:

$$g(f) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right) \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} f^{\frac{n_1-2}{2}} \left(1 + \frac{n_1}{n_2} f\right)^{-\frac{n_1+n_2}{2}}$$

Hình 26 cho biết đồ thị của mật độ phân phối  $F_{15; 5}$  và  $F_{15; 50}$ .



**Hình 26**

Bây giờ giả sử có các mẫu  $(X_1, \dots, X_{n_1})$  từ đại lượng ngẫu nhiên  $X$ ,  $(Y_1, \dots, Y_{n_2})$  từ đại lượng ngẫu nhiên  $Y$  và :

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

$$\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i ; \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$



**Định lý 5.** Nếu  $X$  và  $Y$  độc lập và đều có phân phối chuẩn với cùng phương sai ( $DX=DY$ ) thì  $\frac{S_1^2}{S_2^2}$  có phân phối  $F$  với  $(n_1-1)$  và  $(n_2-1)$  bậc tự do.

#### 4.4. PHÂN PHỐI TIỆM CẬN CHUẨN CỦA ĐẠI LƯỢNG THỐNG KÊ

Theo các định lý giới hạn, khi số lượng của mẫu  $n$  tăng lên vô cùng thì có thể chứng minh được nhiều đại lượng thống kê có phân phối xác suất tiến tới phân phối chuẩn, các phân phối đó được gọi là phân phối tiệm cận chuẩn.

Trong thực tế, với  $n$  khá lớn sự khác nhau giữa phân phối chuẩn và phân phối tiệm cận chuẩn được coi như không đáng kể. Điều này rất quan trọng vì nó cho phép áp dụng những kết quả tốt đối với phân phối chuẩn vào trường hợp phân phối tiệm cận chuẩn.

**4.4.1.** Giả sử có mẫu  $(X_1, \dots, X_n)$  từ đại lượng ngẫu nhiên  $X$  có phân phối xác suất bất kỳ.

**Định lý 1.** Nếu tồn tại  $E(X)=a$  và  $D(X)=\sigma^2$  với  $a$  và  $\sigma$  là hai hằng số thì  $\frac{\sqrt{n}}{\sigma}(\bar{X} - a)$  có phân phối tiến tới phân phối chuẩn  $N(0;1)$ .

**4.4.2.** Ở 4.3. chúng ta đã biết nếu các  $X_i$  ( $i=1, \dots, n$ ) độc lập với nhau, cùng có phân phối chuẩn  $N(0;1)$ , thì  $U = \sum_{i=1}^n X_i^2$  có phân phối xác suất  $\chi^2$  với  $n$  bậc tự do.

**Định lý 2.** Nếu  $n \rightarrow +\infty$  và  $U$  có phân phối  $\chi^2$  với  $n$  bậc tự do thì :

$$\left( \sqrt{2U} - \sqrt{2n-1} \right)$$

có phân phối tiến tới phân phối chuẩn  $N(0;1)$ .

Ở đây tốc độ tiến gần nhanh nên với  $n > 30$  đã có thể coi phân phối của  $\left( \sqrt{2U} - \sqrt{2n-1} \right)$  như phân phối chuẩn  $N(0;1)$ .

**4.4.3.** Giả sử các  $X_i$  ( $i=1, \dots, n$ ) độc lập với nhau và có cùng phân phối chuẩn  $N(\mu; \sigma^2)$  thì như đã biết:

$$t = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

có phân phối Student với  $(n-1)$  bậc tự do:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$

**Định lý 3.** Nếu  $n \rightarrow +\infty$  và phân phối Student với  $(n-1)$  bậc tự do thì phân phối xác suất của  $t$  tiến tới phân phối chuẩn  $N(0;1)$ .

Với  $n > 30$  bảng phân phối Student được coi như trùng với bảng phân phối chuẩn  $N(0;1)$ .

## 4.5. HÀM ƯỚC LƯỢNG CỦA MỘT THAM SỐ CHƯA BIẾT

Trước hết ta xét một vấn đề thực tế như sau:

Cho hai điểm A và B trên mặt đất và không trùng nhau. Để tìm khoảng cách chính xác  $\theta$  giữa A và B người ta tiến hành đo đạc. Giả sử đã đo  $n$  lần và được các kết quả là  $X_1, \dots, X_n$ . Rõ ràng là các  $X_i$  ( $i=1, \dots, n$ ) nói chung khác  $\theta$  vì có sai số, với các số liệu đó người ta không thể tính chính xác được  $\theta$  tuy biết chắc nó tồn tại và nhận giá trị trong khoảng  $(0, +\infty)$ , thường người ta dùng  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  thay cho  $\theta$ ,

và điều này đã được coi là hiển nhiên trong cuộc sống hàng ngày.

Nếu xét kỹ hơn thì ta thấy ở lần đo thứ  $i$  ta nhận được  $X_i = \theta + \delta_i$  với  $\delta_i$  là sai số đo đạc. Với một số giả thiết mà thực tế chấp nhận được thì các  $\delta_i$  là các đại lượng ngẫu nhiên độc lập cùng có phân phối chuẩn  $N(0, \sigma^2)$  với  $\sigma^2$  đã biết, khi ấy  $X_i = \theta + \delta_i$  cũng có phân phối chuẩn  $N(\theta, \sigma^2)$ . Vì các  $X_i$  độc lập nên có thể coi  $(X_1, \dots, X_n)$  là mẫu lấy từ đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\theta, \sigma^2)$  và ta cần tìm hàm ước lượng của  $\theta$  căn cứ vào  $(X_1, \dots, X_n)$ .



Bài toán tìm hàm ước lượng của tham số, một cách tổng quát thường được đặt ra như sau:

Cho  $X$  là đại lượng ngẫu nhiên có luật phân phối  $P(x, \theta)$ , dạng của  $P$  đã biết (do áp dụng lý thuyết chung, do giả thiết hoặc một lý do nào đấy) song  $\theta$  chưa biết và cần tìm  $\theta$ . Thực ra người ta chỉ biết  $\theta$  thuộc một miền  $\Theta$  nào đó đã biết. Việc tìm giá trị thực của tham số  $\theta$  rất khó khăn nên người ta chỉ ước lượng  $\theta$  căn cứ theo kết quả của mẫu. Muốn vậy người ta lấy mẫu  $(X_1, \dots, X_n)$  và lập một đại lượng thống kê  $\hat{\theta}(X_1, \dots, X_n)$  để dùng thay cho  $\theta$ .

### **Định nghĩa 1.**

Đại lượng thống kê  $\hat{\theta}(X_1, \dots, X_n)$  được chọn để dùng thay cho  $\theta$  được gọi là hàm ước lượng của  $\theta$  (hay còn gọi là ước lượng của  $\theta$ ).

$\hat{\theta}(X_1, \dots, X_n)$  là đại lượng ngẫu nhiên vì nó là đại lượng thống kê. Với mỗi giá trị cụ thể  $(X_1, \dots, X_n)$  của mẫu thì  $\hat{\theta}(X_1, \dots, X_n)$  là một điểm trên trục số thực, điểm ấy được dùng thay cho  $\theta$ , vì thế nên còn gọi  $\hat{\theta}(X_1, \dots, X_n)$  là ước lượng điểm của  $\theta$ . Cần chú ý rằng  $\hat{\theta}$  chỉ phụ thuộc  $(X_1, \dots, X_n)$  mà không phụ thuộc  $\theta$ .

*Ví dụ:* Giả sử  $EX = \mu$ ,  $DX = \sigma^2$  và mẫu là  $(X_1, \dots, X_n)$  thì có thể coi  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  là một ước lượng của  $\mu$  và  $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  là ước lượng của  $\sigma^2$ . Trong bài toán mở đầu ta đã lấy  $\bar{X}$  làm hàm ước lượng của  $\theta = EX$ .

Bài toán tìm hàm ước lượng nói trên được gọi là bài toán ước lượng tham số, đó là một trong các bài toán cơ bản của thống kê toán học.

Ứng với một tham số  $\theta$  có vô số hàm ước lượng khác nhau. Vấn đề là phải chọn  $\hat{\theta}(X_1, \dots, X_n)$  theo các tiêu chuẩn nào và thế nào là ước lượng tốt nhất? Từ đó có các định nghĩa sau:

**Định nghĩa 2.**

Hàm ước lượng  $\hat{\theta}(X_1, \dots, X_n)$  của  $\theta$  được gọi là ước lượng không chệch nếu:

$$E \hat{\theta}(X_1, \dots, X_n) = \theta \quad (10)$$

với bất kỳ  $\theta \in \Theta$ .

(Trong công thức (1) kỳ vọng của  $\hat{\theta}(X_1, \dots, X_n)$  được tính theo phân phối xác suất của  $(X_1, \dots, X_n)$  ứng với giá trị của tham số là  $\theta$ . Đối với các công thức tính kỳ vọng về sau ta cũng hiểu với nội dung như vậy).

Nếu coi  $\hat{\theta}(X_1, \dots, X_n) - \theta$  là sai số của ước lượng thì điều kiện (10) chứng tỏ rằng kỳ vọng của sai số bằng không. Nói một cách khác không có sai lầm hệ thống lệch về một phía.

*Ví dụ 1:* Kỳ vọng mẫu  $\bar{X}$  là ước lượng không chệch của  $EX = \mu$ . Thực vậy, vì:

$$E\bar{X} = E \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n EX_i = \mu \quad (11)$$

*Ví dụ 2:* Giả sử  $DX = \sigma^2$  và  $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ , sẽ chứng minh

$\hat{S}^2$  là ước lượng chệch của  $\sigma^2$ :

$$\begin{aligned} E\hat{S}^2 &= \frac{1}{n} E \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} E \sum_{i=1}^n [(X_i - EX_i) - (\bar{X} - E\bar{X})]^2 \\ &= \frac{1}{n} E \sum_{i=1}^n \left\{ (X_i - EX_i)^2 + (\bar{X} - E\bar{X})^2 - 2(X_i - EX_i)(\bar{X} - E\bar{X}) \right\} \\ &= \frac{1}{n} E \sum_{i=1}^n (X_i - EX_i)^2 + \frac{1}{n} E \left\{ n(\bar{X} - E\bar{X})^2 - 2n(\bar{X} - E\bar{X})^2 \right\} \\ &= \frac{n-1}{n} \sigma^2 \end{aligned} \quad (12)$$



Vậy  $\hat{S}^2$  là ước lượng chệch của  $\sigma^2$ . Từ kết quả trên có thể suy ra ngay được rằng:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (13)$$

là ước lượng không chệch của  $\sigma^2$ . Thực ra  $\frac{n-1}{n}$  tiến tới 1 rất nhanh khi  $n$  tăng nên trong thực tế nếu  $n > 50$  thì người ta coi  $\hat{S}^2$  cũng như  $S^2$ .

Độ chính xác của ước lượng không chệch:

Giả sử  $\hat{\theta}(X_1, \dots, X_n)$  là ước lượng không chệch của  $\theta$  và  $D\hat{\theta}(X_1, \dots, X_n) = \sigma^2$ , thì theo bất đẳng thức Trêbưsép ta có:

$$P\left\{\left|\hat{\theta}(X_1, \dots, X_n) - \theta\right| < \varepsilon\sigma\right\} \geq 1 - \frac{\sigma^2}{\varepsilon^2\sigma^2} \geq 1 - \frac{1}{\varepsilon^2} \quad (14)$$

Nếu chọn  $\varepsilon = 3$  thì:

$$P\left\{\left|\hat{\theta}(X_1, \dots, X_n) - \theta\right| < 3\sigma\right\} \geq 1 - \frac{1}{9} \approx 0,889 \quad (15)$$

Công thức (15) đúng với bất kỳ phân phối xác suất của  $X$ . Nếu  $\hat{\theta}(X_1, \dots, X_n)$  có phân phối chuẩn thì vế phải của (15) sẽ lớn hơn và xấp xỉ 0,997.

Trong thực tế người ta thường viết:

$$\left|\hat{\theta}(X_1, \dots, X_n) - \theta\right| < 3\sigma \quad (16)$$

(16) được gọi là công thức  $3\sigma$ . Ta hiểu (16) theo nội dung của (15).

### **Định nghĩa 3.**

Hàm ước lượng  $\hat{\theta}(X_1, \dots, X_n)$  của tham số  $\theta$  được gọi là ước lượng vững nếu với bất kỳ số  $\varepsilon > 0$  cho trước và với bất kỳ  $\theta \in \Theta$ :

$$\lim_{n \rightarrow \infty} P\left\{\left|\hat{\theta}(X_1, \dots, X_n) - \theta\right| < \varepsilon\right\} = 1 \quad (17)$$

ở đây xác suất  $P$  được tính ứng với tham số  $\theta$ .

Điều kiện (17) chỉ rằng: sự kiện  $S_n$  trên đó  $\hat{\theta}(X_1, \dots, X_n)$  khá gần  $\theta$  có xác suất tiến tới 1 ( $P(S_n) \rightarrow +1$ ) khi  $n \rightarrow \infty$ . Ở đây  $S_n$  biến đổi khi  $n$  thay đổi.

Nói cách khác: khi  $n$  khá lớn thì xác suất để sai số tuyệt đối của ước lượng vững không vượt quá  $\varepsilon$  gần bằng 1.

**Định lý 1.** Nếu  $\hat{\theta}(X_1, \dots, X_n)$  là hàm ước lượng của  $\theta$  sao cho:

a)  $\hat{\theta}(X_1, \dots, X_n)$  là ước lượng không chệch của  $\theta$  hoặc  $\lim_{n \rightarrow +\infty} [E\hat{\theta}(X_1, \dots, X_n) - \theta] = 0$  (độ chệch tiến tới 0),

b)  $\lim_{n \rightarrow +\infty} D(X_1, \dots, X_n) = 0$  thì  $\hat{\theta}(X_1, \dots, X_n)$  là ước lượng vững của  $\theta$ .

*Ví dụ 3:* Xét xem  $\bar{X}$  có phải ước lượng vững của  $EX_i = \mu$  hay không?

Trong ví dụ 1 đã chứng minh  $\bar{X}$  là ước lượng không chệch của  $\mu$  vậy điều kiện a) của định lý 1 đã được thỏa mãn, ta xét tiếp:

$$D\bar{X} = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n DX_i = \frac{\sigma^2}{n} \quad (18)$$

để có (18) ta chú ý rằng vì các  $X_i$  độc lập với nhau nên  $\text{cov}(X_i, X_j) = 0$  với  $i \neq j$ . Rõ ràng là  $D\bar{X} = \frac{\sigma^2}{n} \rightarrow 0$  khi  $n \rightarrow +\infty$  và điều kiện b) của

định lý 1 cũng được thỏa mãn, vậy  $\bar{X}$  là ước lượng vững của  $\mu$ .

**Định nghĩa 4:**

Nếu  $\hat{\theta}(X_1, \dots, X_n)$  là ước lượng không chệch của  $\theta$  và  $D\hat{\theta}(X_1, \dots, X_n)$  không lớn hơn phương sai của bất kỳ hàm ước lượng không chệch khác, thì  $\hat{\theta}(X_1, \dots, X_n)$  được gọi là ước lượng không chệch có phương sai bé nhất của  $\theta$ , hay còn gọi là ước lượng hiệu quả của  $\theta$ .



Ta đã biết  $D\hat{\theta}(X_1, \dots, X_n)$  chỉ độ phân tán các giá trị của  $\hat{\theta}$  quanh trị số đúng  $\theta$  (vì  $\hat{\theta}$  không chệch nên  $D\hat{\theta}(X_1, \dots, X_n)$  càng bé thì xác suất để  $\hat{\theta}(X_1, \dots, X_n)$  lấy các giá trị gần  $\theta$  càng lớn;

Một vấn đề được đặt ra là có hàm ước lượng không chệch nào có phương sai bằng 0 hay không? Trả lời vấn đề này có định lý sau:

**Định lý 2.** Cho mẫu  $(X_1, \dots, X_n)$  lấy từ đại lượng ngẫu nhiên  $X$  có mật độ  $f(x, \theta)$  thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế) và  $\hat{\theta}(X_1, \dots, X_n)$  là một hàm ước lượng không chệch bất kỳ của  $\theta$  thì:

$$D\hat{\theta}(X_1, \dots, X_n) \geq \frac{1}{n \cdot E \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2} \quad (19)$$

(19) được gọi là bất đẳng thức thông tin, hay còn gọi là bất đẳng thức Crame-Rao. Công thức (19) khẳng định rằng không tồn tại một hàm ước lượng nào có sai số bằng không và còn cho biết một cận dưới của các phương sai là:

$$\frac{1}{n \cdot E \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2}$$

**Ví dụ 4:** Xét xem  $\bar{X}$  có phải ước lượng hiệu quả của  $EX = \mu$  trong trường hợp  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$  hay không?

Biết rằng  $\frac{\partial \ln f(x, \theta)}{\partial \theta} = \frac{x - \mu}{\sigma^2}$ .

## TÓM TẮT KẾT QUẢ

Tham số cần ước lượng	Hàm ước lượng $\hat{\theta}(X_1, \dots, X_n)$	$E\hat{\theta}(X_1, \dots, X_n)$	$D\hat{\theta}(X_1, \dots, X_n)$	Các tính chất của $\hat{\theta}(X_1, \dots, X_n)$
Kỳ vọng $EX = \mu$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$\mu$	$\frac{\sigma^2}{n}$	Không chệch, vững, hiệu quả (nếu X có phân phối)
Xác suất $p$	Tần suất $\frac{m}{n}$	$p$	$\frac{p(1-p)}{n}$	Không chệch, vững, hiệu quả
Phương sai $\sigma^2$	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	$\sigma^2$	$\frac{1}{n} \left( \mu_4 - \frac{n-3}{n-1} \sigma^4 \right)$ Với $\mu_4 = E(X - EX)^4$	Không chệch, vững

Vậy:

$$n \cdot E \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 = n E \frac{(X - \mu)^2}{\sigma^4} = \frac{n}{\sigma^2}$$

Ta biết  $D\bar{X} = \frac{\sigma^2}{n}$ , nghĩa là  $D\bar{X}$  bằng biểu thức ở vế phải của (19), vậy  $\bar{X}$  là ước lượng hiệu quả của  $\mu$ .

Ở đây coi tần suất  $\frac{m}{n}$  là đại lượng ngẫu nhiên và  $\frac{m}{n} = \frac{\sum_{i=1}^n X_i}{n}$  với  $X_i$  ( $i=1, \dots, n$ ) là các đại lượng ngẫu nhiên độc lập:  $X_i$  nhận giá trị 1 với xác suất  $p$  và giá trị 0 với xác suất  $(1-p)$ , vì vậy có thể tính  $E\left(\frac{m}{n}\right)$  và  $D\left(\frac{m}{n}\right)$  dẫn đến kết quả như trên.



*Ví dụ 5:* Trong một xí nghiệp để tính số đơn vị nguyên liệu cần thiết để sản xuất ra một thành phẩm người ta lấy mẫu với số lượng 20 và được các số liệu: 3,0; 3,8; 3,1; 3,2; 3,5; 3,2; 3,5; 3,6; 3,3; 3,8; 3,5; 3,2; 4,0; 3,6; 3,4; 3,5; 4,3; 3,5; 3,0; 4,0. Hỏi số lượng đơn vị nguyên liệu cần thiết cho mỗi thành phẩm là bao nhiêu?

Gọi  $X$  là đại lượng ngẫu nhiên chỉ số lượng đơn vị nguyên liệu cần thiết để sản xuất ra một thành phẩm, như vậy cần ước lượng  $EX = \mu$ .

Căn cứ theo mẫu, lấy  $\bar{X}$  làm ước lượng cho  $\mu$  (ở đây

$\bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i = 3,5$  và  $\frac{S}{\sqrt{n}} = 0,08$ ) ta có:  $\mu = 3,5 \pm 0,24(3 \times 0,08)$  với xác suất là 0,889.

*Ví dụ 6:* Trong số dân của thành phố Vácsava chọn hủ hoá 900 người, sau khi hỏi từng người xem hè vừa qua nghỉ ở đâu thì biết có 200 người nghỉ ở miền núi. Vậy kết luận thế nào về tỷ số người Vácsava nghỉ hè ở miền núi nói chung?

Ở đây cần ước lượng xác suất  $p$  của sự kiện (người Vácsava nghỉ hè ở miền núi). Ta dùng tần suất làm ước lượng cho xác suất.

Lập tần suất  $\frac{m}{n} = \frac{200}{900} = 0,222$  có thể nói ngay có 22,2% dân số

Vácsava nghỉ hè ở miền núi nói chung, tuy nhiên ta cần biết sai số trung bình là bao nhiêu.

Biết rằng  $D\left(\frac{m}{n}\right) = \frac{p(1-p)}{n}$  nhưng  $p$  không biết nên phải thay bằng  $\frac{m}{n}$  và  $D\left(\frac{m}{n}\right) \approx \frac{2/9(1-2/9)}{900} \approx (0,013)^2$ , từ đó ta có tỷ số cần tìm là:

$$0,222 \pm 3 \times 0,013 \approx 0,22 \pm 0,04.$$

## 4.6 ƯỚC LƯỢNG THAM SỐ THEO PHƯƠNG PHÁP HỢP LÝ CỰC ĐẠI

Ước lượng tham số theo phương pháp hợp lý cực đại do Phiso đề ra là một trong những phương pháp ước lượng quan trọng và hay dùng nhất.

Giả sử đại lượng ngẫu nhiên  $X$  có mật độ phân phối xác suất  $f(x, \theta)$ , với dạng của  $f$  đã biết còn  $\theta$  chưa biết. Để ước lượng  $\theta$ , ta lấy mẫu  $(X_1, \dots, X_n)$  và lập hàm số:

$$L(\theta) = f(X_1, \theta) \times \dots \times f(X_n, \theta) \quad (20)$$

$L(\theta)$  được gọi là hàm hợp lý của mẫu, nó phụ thuộc  $(X_1, \dots, X_n)$  và  $\theta$  nhưng ta coi  $(X_1, \dots, X_n)$  là các hằng số còn  $\theta$  được coi là biến số.

Coi  $(X_1, \dots, X_n)$  là các hằng số vì trong thực tế sau khi lấy mẫu thì ta đã biết các trị số cụ thể của  $(X_1, \dots, X_n)$  là các hằng số. Vấn đề đặt ra là tìm  $\hat{\theta}(X_1, \dots, X_n)$  sao cho:

$$L(\hat{\theta}(X_1, \dots, X_n)) \geq L(\theta) \quad (21)$$

với bất kỳ  $\theta \in \Theta$  ( $\Theta$  là miền giá trị của  $\theta$ , nói cách khác  $\Theta$  là tập hợp các trị số mà  $\theta$  có thể nhận được). Vì hàm lôgarit đơn điệu nên ta còn xác định  $\hat{\theta}(X_1, \dots, X_n)$  từ:

$$l(\hat{\theta}(X_1, \dots, X_n)) \geq l(\theta) \quad (22)$$

với bất kỳ  $\theta \in \Theta$ , và với:

$$l(\hat{\theta}) = \log_e L(\theta) = \ln L(\theta)$$

Ước lượng  $\hat{\theta}(X_1, \dots, X_n)$  được xác định bởi (21) hoặc (22) được gọi là ước lượng hợp lý cực đại của  $\theta$ .

Nếu điều kiện có đạo hàm được thỏa mãn thì tại  $\hat{\theta}(X_1, \dots, X_n)$ :

$$\frac{dl(\theta)}{d\theta} = 0 \quad (23)$$

Phương trình (23) được gọi là phương trình hợp lý và mọi nghiệm của nó, nếu thỏa mãn (21) hoặc (22) đều là ước lượng hợp lý cực đại



của  $\theta$ . Thường người ta tìm ước lượng hợp lý cực đại dưới dạng nghiệm của (23).

*Ví dụ 1:* Giả sử  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$ ,  $\sigma^2$  đã biết, và mẫu có được là  $(X_1, \dots, X_n)$ . Hãy tìm ước lượng hợp lý cực đại của  $\mu$ .

Trước hết ta lập  $L(\theta)$

$$L(\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2}$$

ở đây  $\theta = \mu$ .

$$\ell(\theta) = \ln L(\theta) = n \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{d\ell(\mu)}{d\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)$$

vậy phương trình hợp lý là:

$$\sum_{i=1}^n (X_i - \mu) = 0$$

Giải phương trình này ta được ước lượng hợp lý cực đại của  $\mu$  là:

$$\hat{\mu}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

(vì  $\frac{d^2\ell(\theta)}{d\theta^2} = -\frac{n}{\sigma^2} < 0$  nên tại  $\hat{\mu}$  hàm hợp lý đạt giá trị lớn nhất).

Trước đây đã giả thiết  $\theta$  là một tham số ( $\theta$  lấy các giá trị trên trục số thực) tuy nhiên các kết quả trên có thể mở rộng cho trường hợp  $\theta$  là một đại lượng nhiều chiều chẳng hạn  $\theta = (\theta_1, \theta_2)$ , khi ấy ứng với (23) có hệ phương trình hợp lý:

$$\frac{\partial \ell(\theta)}{\partial \theta_i} = 0, \quad (i=1,2) \quad (24)$$

Nghiệm của (24) có dạng

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$$

cũng được gọi là ước lượng hợp lý cực đại của  $\theta$  nếu thỏa mãn (21) hoặc (22).

**Ví dụ 2:** Giả sử  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$  trong đó  $\mu$  và  $\sigma^2$  đều chưa biết. Với mẫu  $(X_1, \dots, X_n)$  hãy tìm ước lượng hợp lý cực đại của  $\mu$  và  $\sigma^2$ .

Ở đây có  $\theta = (\mu, \sigma^2)$  và hệ các phương trình hợp lý là:

$$\frac{\partial \ell(\theta)}{\partial \mu} = \sum_{i=1}^n X_i - n\mu$$

$$\frac{\partial \ell(\theta)}{\partial \sigma^2} = \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2}$$

với  $\ell(\theta)$  xác định như ở ví dụ 1.

Giải ra ta được:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Bây giờ cần thử lại xem tại  $\left( \hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right)$

(22) có được thỏa mãn hay không?

Biết rằng:

$$\ell(\theta) = n \ln \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$$

cho nên :

$$\ell(\hat{\mu}, \hat{\sigma}^2) = -n \ln \hat{\sigma} - \frac{n}{2}$$



Có thể viết:

$$l(\mu, \sigma^2) = -n \ln \sigma - \frac{n\hat{\sigma}^2 + n(\bar{X} - \mu)^2}{2\sigma^2}$$

vì:

$$l(\mu, \sigma^2) \leq l(\hat{\mu}, \hat{\sigma}^2)$$

nghĩa là:

$$-n \ln \sigma - \frac{n\hat{\sigma}^2 + n(\bar{X} - \mu)^2}{2\sigma^2} \leq -n \ln \hat{\sigma} - \frac{n}{2}$$

hay:

$$0 \leq \left\{ \left( \frac{\hat{\sigma}^2}{2\sigma^2} - \frac{1}{2} \right) - \ln \frac{\hat{\sigma}}{\sigma} \right\} + \frac{(\bar{X} - \mu)^2}{2\sigma^2}$$

và điều đó đúng vì với  $\forall x > 0$ :

$$\ln x < \left( x^2 - 1 \right) \frac{1}{2}$$

Vậy  $\bar{X}$  và  $\hat{\sigma}^2$  là ước lượng hợp lý cực đại của  $\mu$  và  $\sigma^2$ .

Trường hợp đại lượng ngẫu nhiên  $X$  có phân phối rời rạc, người ta cũng định nghĩa ước lượng hợp lý cực đại tương tự như trên.

Ước lượng hợp lý cực đại có nhiều tính chất tối ưu, ví dụ nó là ước lượng vững, với  $n$  khá lớn thì nó có phân phối gần với phân phối chuẩn và nó cũng khá gần với ước lượng hiệu quả.

*Chú ý 1.* Khi định nghĩa ước lượng hợp lý cực đại bởi (21) hoặc (22) người ta dựa theo quan điểm: "giá trị đúng của  $\theta$  trong thực tế chính là giá trị ứng với xác suất xảy ra lớn nhất". Vì thế ước lượng ứng với xác suất lớn nhất được coi là hợp lý nhất.

*Chú ý 2.* Ước lượng được định nghĩa bởi (21) hoặc (22) có thể được gọi là ước lượng hợp lý cực đại để bạn đọc thấy sự liên hệ với ước lượng là nghiệm của (23).

## 4.7. ƯỚC LƯỢNG THAM SỐ BẰNG KHOẢNG TIN CẬY

Ở đây ta xét một phương pháp tìm  $\theta$  khác với phương pháp hàm ước lượng ở 4.5.

**Định nghĩa.** Khoảng tin cậy  $(\hat{\theta}_1, \hat{\theta}_2)$  của tham số  $\theta$  với độ tin cậy  $Y$  là một khoảng với hai đầu mút:

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n); \hat{\theta}_2 = \hat{\theta}_2(X_1, \dots, X_n)$$

và 
$$P = \left\{ \hat{\theta}_1(X_1, \dots, X_n) \leq \theta \leq \hat{\theta}_2(X_1, \dots, X_n) \right\} = Y \quad (25)$$

Bài toán tìm khoảng tin cậy của  $\theta$  thường đặt ra như sau: Cho trước hằng số  $Y$  hãy xác định  $\hat{\theta}_1(X_1, \dots, X_n)$  và  $\hat{\theta}_2(X_1, \dots, X_n)$  thỏa mãn (25).

Theo định nghĩa ta thấy  $Y$  càng lớn và độ dài khoảng  $(\hat{\theta}_1, \hat{\theta}_2)$  càng nhỏ thì ước lượng càng chính xác.

Sau đây ta xét vài ví dụ:

*Ví dụ 1:* Đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$  với  $\sigma^2$  đã biết,  $\mu$  chưa biết. Để ước lượng  $\mu$  ta lấy mẫu  $(X_1, \dots, X_n)$  và dùng phương pháp khoảng tin cậy như sau:

Xét đại lượng thống kê phụ  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  có phân phối chuẩn  $N(0,1)$ , theo bảng của hàm số Laplat  $\phi(u)$  có thể tìm được một số  $u_\gamma > 0$  sao cho:

$$P\left\{ -u_\gamma \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq u_\gamma \right\} = \gamma \quad (26)$$

Giả sử  $\gamma = 95\%$ , theo bảng II (phần phụ lục) ta có  $u_\gamma = 1,96$ , nếu  $\gamma = 99\%$  thì  $u_\gamma = 2,58$ ...

Vậy khoảng tin cậy của  $\mu$  được cho bởi:

$$P\left\{ \bar{X} - u_\gamma \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_\gamma \frac{\sigma}{\sqrt{n}} \right\} = \gamma \quad (27)$$



Ví dụ 2: Xét bài toán ở ví dụ 1 với trường hợp  $\sigma$  chưa biết.  
Để ước lượng  $\mu$  bây giờ phải dùng đại lượng thống kê phụ:

$$\frac{\bar{X} - \mu}{S} \sqrt{n}$$

(  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, S = \sqrt{S^2}$  ). Đại lượng này có phân phối

Stiudon với  $(n-1)$  bậc tự do, từ đây ta có công thức:

$$P\left\{-t_{n,\gamma} \leq \frac{\bar{X} - \mu}{S} \sqrt{n} \leq t_{n,\gamma}\right\} = \gamma \quad (28)$$

$t_{n,\gamma}$  xác định theo bảng VII (phần phụ lục), từ (28) suy ra:

$$P\left\{\bar{X} - t_{n,\gamma} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n,\gamma} \frac{S}{\sqrt{n}}\right\} = \gamma \quad (29)$$

Công thức (29) cho khoảng tin cậy cần tìm của  $\mu$ . Nếu  $n$  khá lớn thì các công thức (27) và (29) cho các kết quả gần như nhau.

Ví dụ 3: Một giống lúa mới được gieo trong 10 miếng đất thí nghiệm có các điều kiện giống nhau, cho các sản lượng tính theo cùng đơn vị như sau:

25,4; 28,0; 20,1; 27,4; 25,6; 23,9; 24,8; 26,4; 27,0; 25,4.

Hãy tìm khoảng tin cậy của sản lượng giống lúa trên với độ tin cậy 95%.

Ở đây coi sản lượng lúa là đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$  với  $\mu$  là tham số chưa biết cần ước lượng và  $\sigma^2$  cũng chưa biết. Ta tính ngay được  $\bar{X} = 25,4$ ;  $S = 2,27$ .

Theo bảng VII (phần phụ lục), với  $\gamma = 0,95$  ta có  $t_{n-1,\gamma} = t_{9;0,95} = 2,26$ . Vậy theo (29):

$$P\left\{25,4 - 2,26 \cdot \frac{2,27}{\sqrt{10}} \leq \mu \leq 25,4 + 2,26 \frac{2,27}{\sqrt{10}}\right\} = 0,95$$

hay

$$P\{23,8 \leq \mu \leq 27\} = 0,95.$$

Để ước lượng khoảng tin cậy của  $\sigma^2$  với độ tin cậy  $\gamma$  ta ứng dụng định lý 3 về phân phối  $\chi^2$  (4.3) theo đó nếu  $X$  ở tập hợp toàn bộ có phân phối chuẩn, thì  $\frac{n-1}{\sigma^2} S^2$  có phân phối  $\chi^2$  với  $(n-1)$  bậc tự do. Theo bảng phân phối của  $\chi^2$  ta sẽ tìm được 2 số  $u_1$  và  $u_2$  sao cho:

$$P\left(u_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq u_2\right) = \gamma \quad (30)$$

Các số  $u_1$  và  $u_2$  như vậy có rất nhiều, vì thế thường chọn  $u_1$  và  $u_2$  sao cho:

$$P(\chi^2 < u_1) = P(\chi^2 > u_2) = \frac{1-\gamma}{2} \quad (31)$$

Từ đó ta có:

$$P\left\{\frac{(n-1)S^2}{u_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{u_1}\right\} = \gamma \quad (32)$$

cho khoảng tin cậy cần tìm.

*Ví dụ 4:* Hãy sử dụng các số liệu ở ví dụ 3 để tìm khoảng tin cậy của phương sai sản lượng lúa với độ tin cậy 0,9.

Từ phân phối  $\chi^2$  với 9 bậc tự do và  $\gamma = 0,9$ , theo (31) và bảng VI (phần phụ lục) ta xác định được  $u_1 = 3,325$  và  $u_2 = 16,919$ .

Theo mẫu đã biết thì tính được  $S^2 = (2,27)^2$ , vậy theo (32):

$$P\left\{\frac{9(2,27)^2}{16,919} \leq \sigma^2 \leq \frac{9(2,27)^2}{3,325}\right\} = 0,90$$

hay:

$$P\{2,65 \leq \sigma^2 \leq 13,5\} = 0,90.$$



## 4.8. ƯỚC LƯỢNG THAM SỐ VỚI MẪU CÓ SỐ LƯỢNG LỚN

Với số lượng của mẫu khá lớn thì việc ước lượng tham số trở nên đơn giản hơn, khi ấy các phân phối tiệm cận chuẩn được thay thế bởi phân phối chuẩn tương ứng cùng như các đặc số của phân phối chính xác được thay bởi các đặc số mẫu tương ứng.

*Ví dụ 1:* Xem 4.000 gia đình công nhân thì thấy 3.200 gia đình có máy thu thanh. Hãy tìm khoảng tin cậy của tỷ số gia đình công nhân có máy thu thanh nói chung với độ tin cậy 0,95.

Ở đây phải ước lượng  $p$  (tỷ số gia đình có máy thu thanh ở tập hợp chính, hay nói cách khác  $p$  là xác suất để một gia đình có máy thu thanh) biết tần suất  $\frac{m}{n}$  tương ứng là  $\frac{3.200}{4.000} = 0,8$ . Coi  $\frac{m}{n}$  là đại lượng ngẫu nhiên có phân phối tiệm cận chuẩn  $E\left(\frac{m}{n}\right) = p$  và

$D\left(\frac{m}{n}\right) = \frac{p(1-p)}{n}$ , vì  $n=4.000$  khá lớn nên ta xem  $\frac{m}{n}$  như có phân phối chuẩn với kỳ vọng là  $p$  và phương sai:

$$\frac{0,8 \times (1 - 0,8)}{4.000} = 0,4 \cdot 10^{-4}$$

Khi ấy  $\frac{\frac{m}{n} - p}{\sqrt{0,4 \cdot 10^{-4}}}$  có phân phối chuẩn  $N(0,1)$  và khoảng tin

cậy của  $p$  được xác định từ:

$$\left\{ -u \leq \frac{\frac{m}{n} - p}{\sqrt{0,4 \cdot 10^{-4}}} \leq u \right\} = 0,95$$

theo bảng II (phần phụ lục),  $u_7 = 1,96$ , vậy khoảng tin cậy cần tìm là:  $0,788 \leq p \leq 0,812$ .

*Ví dụ 2:* Để ước lượng sai số trung bình của chiều dài một chi tiết máy sản xuất hàng loạt, người ta lấy mẫu gồm 1.376 phần tử. Sai số

trung bình  $\bar{X} = 70,4$  và  $S^2 = (0,37)^2$ . Hãy tìm khoảng tin cậy  $EX$ , với độ tin cậy  $\gamma = 0,99$ .

Gọi  $EX$  là sai số cần biết, giả sử  $X$  có phân phối chuẩn với kỳ vọng và phương sai đều chưa biết. Nếu mẫu có số lượng bé ta phải dùng phân phối Stiuđon để tính, nhưng ở đây vì  $n$  lớn nên có thể coi đại lượng  $t$  có phân phối chuẩn  $N(0,1)$  (hoặc  $\bar{X}$  có phân phối chuẩn  $N\left(\bar{X}, \frac{S^2}{n}\right)$ ). Theo bảng II (phần phụ lục) thì  $u_{0,99}=2,58$  vậy khoảng tin

cậy cần tìm của  $EX$  là:

$$\left\{ 70,4 - 2,58 \cdot \frac{0,37}{\sqrt{1376}} \leq EX \leq 70,4 + 2,58 \cdot \frac{0,37}{\sqrt{1376}} \right\}$$

hay:  $[70,37 \leq EX \leq 70,43]$

## BÀI TẬP CHƯƠNG 4

**Bài 1:** Khi đo độ dài của 30 chi tiết chọn ngẫu nhiên của một loại sản phẩm, người ta thu được số liệu sau đây:

39, 41, 40, 43, 41, 40, 44, 42, 41, 43, 41, 42, 39, 40, 42, 43, 41, 41, 42, 39, 42, 42, 41, 42, 40, 41, 43, 41, 39, 40.

Xây dựng:

- Bảng phân phối thực nghiệm và hàm phân phối mẫu tương ứng;
- Đường cong gấp khúc phân phối của độ dài chi tiết đã đo.

**Giải:** Số liệu đã cho cần tập hợp lại vào bảng số sau đây:

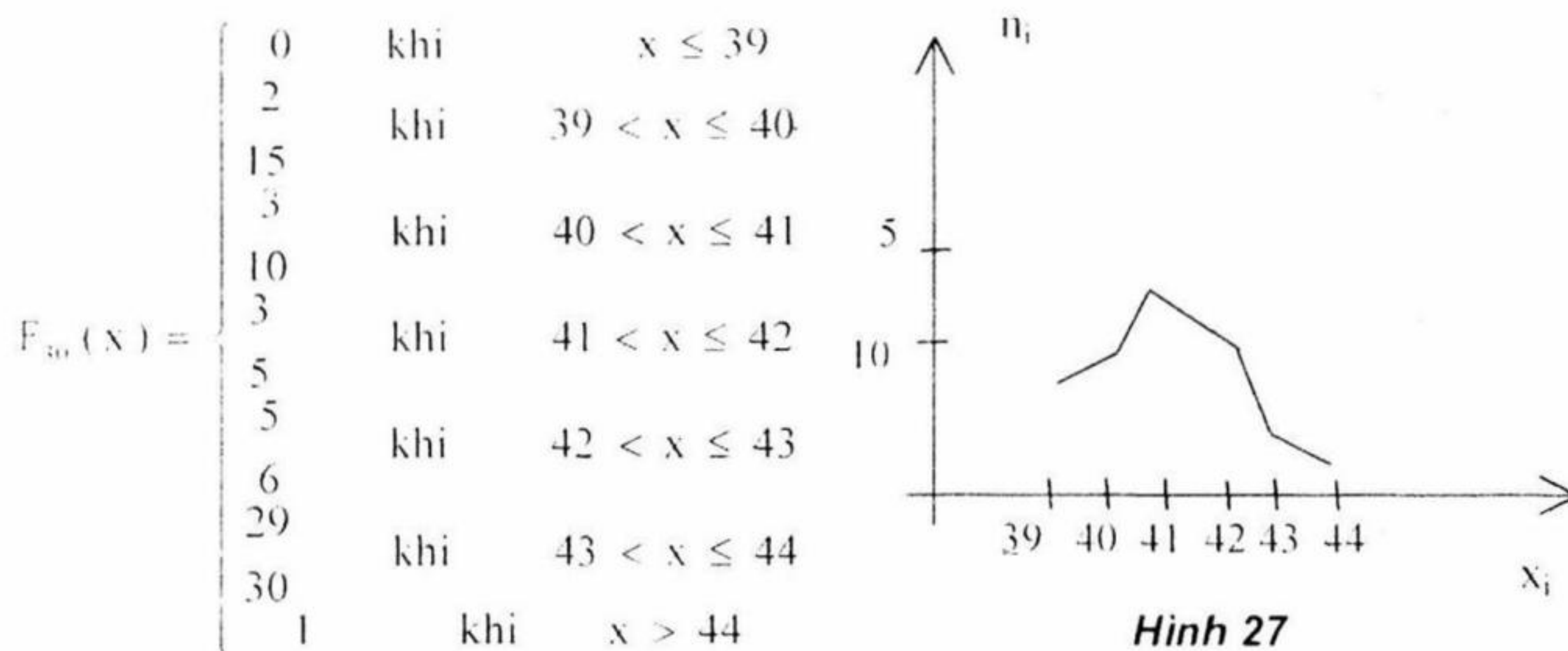
X	39	40	41	42	43	44	
N	4	5	9	7	4	1	$\Sigma = 30$

a) Bảng phân phối thực nghiệm của bộ số liệu gốc sẽ là:

X	39	40	41	42	43	44
W	$\frac{2}{15}$	$\frac{1}{6}$	$\frac{3}{10}$	$\frac{7}{30}$	$\frac{2}{15}$	$\frac{1}{30}$



Từ bảng số trên bằng cách tìm các tần suất tích lũy ta có hàm phân phối mẫu:



Nếu biểu diễn trên đồ thị đây là một hàm bậc thang.

b) Đường gấp khúc phân phối cho bởi hình 27. Ta chọn đơn vị của trục tung là  $n_i$ -tần số của giá trị  $x_i$  tương ứng.

**Bài 2:** Cho bộ số liệu gồm 79 giá trị như sau:

43 52 53 54 45 54 59 54 48 40 56 53 59 42 43  
 49 51 51 44 57 51 45 50 42 51 56 57 50 50 51  
 55 50 53 48 46 47 51 51 57 52 48 56 50 48 53  
 57 55 50 50 49 52 48 51 60 48 53 62 51 49 61  
 42 50 51 60 51 50 57 55 49 53 56 49 46 55 56  
 47 50 47 54.

Hãy xây dựng bảng phân phối xác suất thực nghiệm.

**Giải:** Kết quả cho bởi bảng sau đây: Cột 1 là các nhóm số liệu, cột 2 và 3 là tần số và tần số tích lũy, cột 4 và 5 là tần suất và tần suất tích lũy.

Nhóm	Tần số	Tần số tích lũy	Tần suất	Tần suất tích lũy (%)
40 – < 44	6	6	7,59	7,59
44 – < 48	9	15	11,39	18,99
48 – < 52	31	46	39,24	58,23
52 – < 56	17	63	21,52	79,75
56 – < 60	12	75	15,19	94,94
60 – < 64	4	79	5,06	100,00

Để ý rằng cột 4 cho ta một hình ảnh xấp xỉ về hàm phân phối mẫu. Cách xây dựng bảng phân phối thực nghiệm dưới dạng cột thường rất phổ biến trong các tài liệu của Mỹ-Anh.

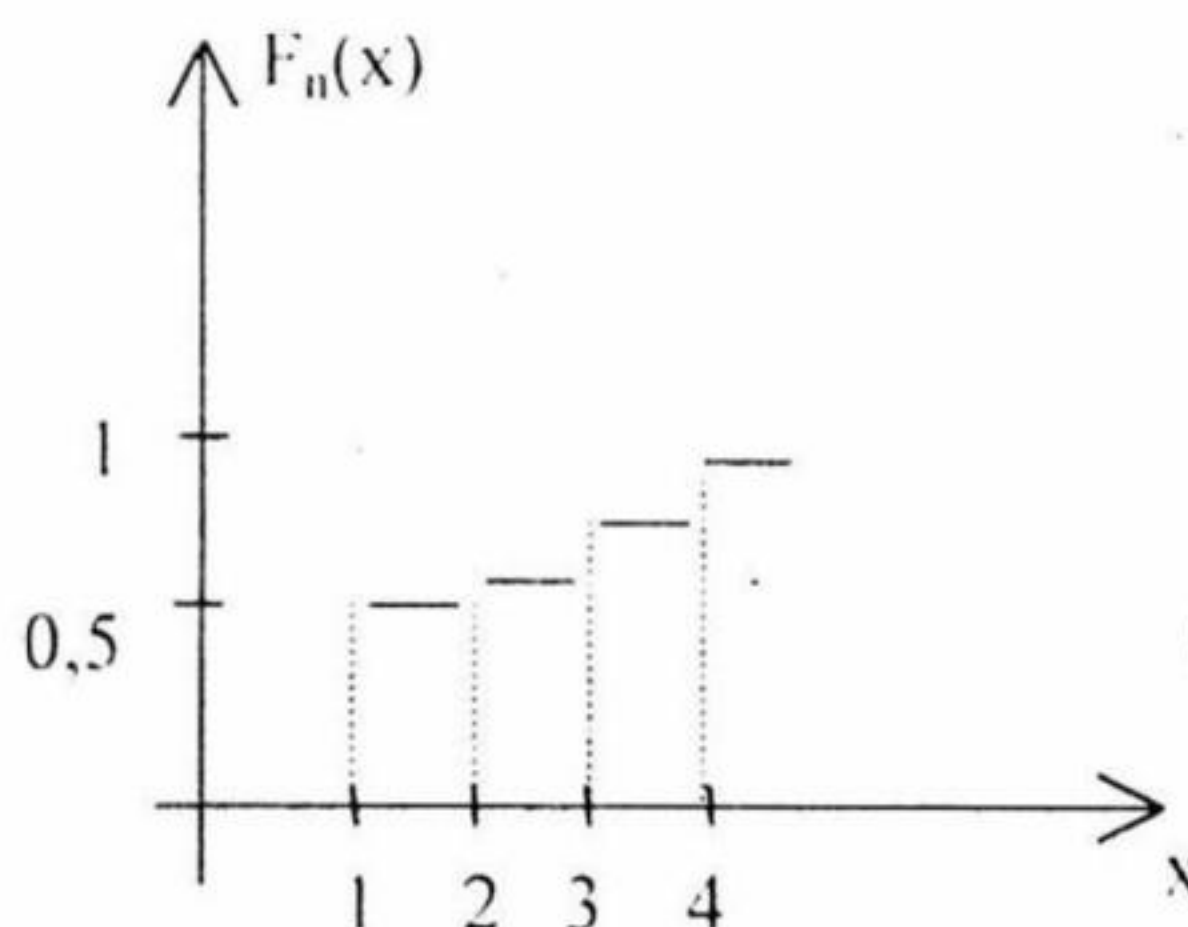
**Bài 3:** Cho một bảng phân phối thực nghiệm như sau:

X	1	2	3	4
W	0,4	0,1	0,3	0,2

Tìm hàm phân phối mẫu và vẽ đồ thị của nó.

**Giải:** Ta có hàm phân phối mẫu:

$$F_n(x) = \begin{cases} 0 & \text{khi } x \leq 1, \\ 0,4 & \text{khi } 1 < x \leq 2, \\ 0,5 & \text{khi } 2 < x \leq 3, \\ 0,8 & \text{khi } 3 < x \leq 4 \\ 1 & \text{khi } x > 4 \end{cases}$$



**Hình 28**

Đồ thị của  $F_n(x)$  biểu diễn trên hình 28.



**Bài 4:** Cho biến ngẫu nhiên  $X$  có phân phối  $N(4,1)$ . Tìm xác suất  $P\left(X > \sqrt{\chi_9^2} + 4\right)$  (ký hiệu  $\chi_9^2$  là biến có phân phối  $\chi^2$  với 9 bậc tự do).

**Giải:** Ta tìm cách biến đổi biểu thức  $X > \sqrt{\chi_9^2} + 4$  về dạng có phân phối xác định (ở đây  $\chi_9^2$  là biến ngẫu nhiên có phân phối  $\chi^2$  với 9 bậc tự do). Để làm điều ấy ta biến đổi bất đẳng thức trên:

$$\frac{1}{3}(X - 4) > \frac{\sqrt{\chi_9^2}}{3}$$

Đặt  $Z = X - 4$ , dễ thấy  $Z \sim N(0,1)$  và:

$$\frac{Z}{3} > \sqrt{\frac{\chi_9^2}{9}},$$

kéo theo  $\frac{Z}{\sqrt{\chi_9^2/9}} > 3$ .

Từ đó xác suất cần tìm có dạng  $P(T > 3)$ , với  $T = \frac{Z}{\sqrt{\chi_9^2/9}} \sim t(9)$ .

Dùng bảng số chi tiết hoặc phần mềm máy tính tương ứng sẽ tìm được  $P(T > 3) = 0,0075$ .

**Bài 5:** Cho  $X$  là biến ngẫu nhiên  $\sim \chi^2(9)$ . Tìm xác suất để  $X$  không vượt quá giá trị 2,7.

**Giải:** Ta phải tìm  $P(X < 2,7)$ . Theo bảng phân phối  $\chi^2$  (xem phụ lục) ở giao điểm của hàng  $n=9$  với cột  $\alpha=0,975$  ta có giá trị 2,7; điều đó có nghĩa là:

$$P(X > 2,7) = 0,975.$$

Từ đó xác suất cần tìm là 0,025.

**Bài 6:** Tìm xác suất để biến ngẫu nhiên  $T \sim t(17)$ , có giá trị:

- Không bé hơn 2,11;
- Không bé hơn -2,11.

**Giải:**

a) Ta phải tìm  $P(T \geq 2,11)$ . Tra bảng phân phối  $t$ , tìm giao điểm của hàng  $n=17$  với cột  $\alpha=0,05$  ta có giá trị 2,11; vậy xác suất cần tìm là  $0,5/2=0,025$  (xem phụ lục).

b) Do tính đối xứng của phân phối  $t$ ,  $P(T \geq -2,11)=1-0,025=0,975$  (và bằng  $P(T < 2,11)$ ). Chú ý rằng  $P(|T| \geq 2,11) = 0,05$  theo bảng Stiuđon (tìm  $t(17;0,05)=2,11$ ).

**Bài 7:** Cường độ chịu lực của một loại vật liệu xây dựng tuân theo phân phối  $\chi^2_{22}$ ; còn áp lực lên loại vật liệu đó sau khi xây dựng có luật  $\chi^2_9$ . Tính xác suất để vật liệu đó chịu được áp lực đã biết.

**Giải:** Trên thực tế ta phải tìm xác suất để cường độ chịu lực của vật liệu phải bé hơn áp lực đã biết, tức là tìm  $P\left(\frac{\chi^2_9}{\chi^2_{22}} < 1\right)$ . Để ý tỷ số

$\left(\frac{\chi^2_9/9}{\chi^2_{22}/22}\right) = F$  sẽ có phân phối  $F(9,22)$ , (xem 4.3-(9)). Tra bảng phân phối  $F$  với  $\alpha=0,05$ , giao điểm của hàng  $n_1=9$  và cột  $n_2=22$  cho ta kết quả  $2,34 \approx \frac{22}{9}$ . Điều đó có nghĩa là:

$$P(F > 2,34) \approx P\left(\frac{\chi^2_9}{\chi^2_{22}} > 1\right) = P\left(\frac{\chi^2_9/9}{\chi^2_{22}/22} > \frac{22}{9}\right) = P(F > 2,33).$$

Suy ra xác suất:

$$P\left(\frac{\chi^2_9}{\chi^2_{22}} > 1\right) \approx 0,05;$$

từ đó xác suất cần tìm sẽ là  $1-0,05 = 0,95$ .

**Bài 8:** Cho mẫu  $X_1, X_2, \dots, X_n$  cảm sinh bởi một biến ngẫu nhiên  $X \sim N(a, \sigma^2)$ . Hãy tìm ước lượng hợp lý nhất của hai tham số  $a$  và  $\sigma^2$ .

**Giải:** Hàm mật độ chuẩn, ký hiệu là  $f(x, a, \sigma^2)$  đã biết ở chương 3. Hàm hợp lý, ký hiệu là  $L(X_1, X_2, \dots, X_n, a, \sigma^2)$ , sẽ có dạng:



$$L = \prod_{i=1}^n f(x_i, a, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2}$$

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2.$$

Chú ý là điểm cực đại của  $\ln L$  cũng chính là điểm cực đại của  $L$ , nên ta có thể giải hệ  $\frac{\partial \ln L}{\partial \theta_i} = 0$  ( $i=1,2$ ).

Hệ phương trình hợp lý sẽ là:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a) = 0, \\ \frac{\partial \ln L}{\partial \theta_2} = \frac{1}{4\sigma^2} \sum_{i=1}^n (x_i - a)^2 - \frac{n}{2\sigma^2} = 0. \end{cases}$$

Rút ra từ phương trình đầu thay vào phương trình thứ hai, ta có thể tìm được các nghiệm một cách dễ dàng (chú ý là nếu hệ có nghiệm duy nhất thì nó chính là điểm cực đại của hàm hợp lý):

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X},$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{a})^2 = S^2.$$

Như vậy phương sai mẫu là ước lượng hợp lý nhất (mặc dù là ước lượng chệch) của phương sai.

**Bài 9:** Độ dày của bản kim loại giả sử là tuân theo luật chuẩn. Đo 10 bản kim loại đó ta thu được số liệu sau:

4,1 3,9 4,7 4,4 4,0 3,8 4,4 4,2 4,4 5,0

Hãy xác định:

- Khoảng tin cậy 90% cho độ dày trung bình trên;
- Khoảng tin cậy 95% cho phương sai của độ dày đó.

**Giải:** Trước hết ta phải tính các ước lượng  $\bar{X}$  và  $S^2$  cho kỳ vọng và phương sai của độ dày đang xét:

$$\bar{X} = \frac{1}{10} (4,1 + 3,9 + \dots + 5,0) = 4,29;$$

$$S^2 = \frac{1}{9} [(4,1 - 4,29)^2 + (3,9 - 4,29)^2 + \dots + (5,0 - 4,29)^2];$$

hoặc dùng công thức:

$$S^2 = \frac{1}{9} \left( \sum X_i^2 - 10(\bar{X})^2 \right) = \frac{1,23}{9} \approx 0,1367 \text{ và } S = 0,37.$$

a) Khoảng tin cậy 90% có dạng (để ý  $\phi(1,645) = 0,45$ ).

$$\left( 4,29 - 1,645 \frac{0,37}{\sqrt{10}}; 4,29 + 1,645 \frac{0,37}{\sqrt{10}} \right) \approx (4,09; 4,49).$$

b) Để tìm khoảng tin cậy 95% cho phương sai, ta phải tìm các giá trị  $\chi^2(9; 0,975)$  và  $\chi^2(9; 0,025)$ . Tra bảng phân phối  $\chi^2$  chúng lần lượt là 2,7 và 19,02. Từ đó ta có khoảng tin cần tìm:

$$\left( \frac{1,23}{19,2}; \frac{1,23}{2,7} \right) \approx (0,064; 0,456).$$

**Bài 10:** Để xác định độ chính xác của một chiếc cân ta không có sai số hệ thống, người ta tiến hành 5 lần cân độc lập (cùng một vật), kết quả như sau:

94,1   94,8   96,0   95,4   95,2   (kg)

Xác định ước lượng không chệch của phương sai số đo trong hai trường hợp:

- Biết khối lượng vật cân là 95;
- Không biết khối lượng vật cân.

**Giải:**

a) Khi đã biết trị trung bình lý thuyết  $a = 95$  thì ước lượng không chệch của phương sai được tính theo công thức:



$$\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 = \frac{1}{5} \sum_{i=1}^5 (X_i - 95)^2 = 0,41$$

b) Nếu không biết  $a$ , ta phải ước lượng theo công thức:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{5} \sum_{i=1}^5 X_i = 95,5$$

Từ đó ước lượng không chệch của phương sai là:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{4} \sum_{i=1}^5 (X_i - 95,5)^2 = 0,7$$

**Bài 11:** Đo lượng huyết tương của 8 người mạnh khỏe ta có:

2,75 2,86 3,37 2,76 2,62 3,49 3,05 3,12 (lít)

Xác định các đặc trưng mẫu:

a) Trung bình;

b) Trung vị;

c) Mốt.

**Giải:**

a) Ở đây  $n = 8$  và việc tìm trung bình mẫu là không có gì khó khăn:

$$\bar{X} = \frac{1}{8} \sum_{i=1}^8 X_i = 3,001$$

b) Để tìm trung vị mẫu, ta sắp xếp lại dãy số liệu:

2,62 2,75 2,76 2,86 3,05 3,12 3,37 3,49

Từ đó trung vị mẫu sẽ là giá trị nằm ở vị trí  $(8+1)/2=4,5$ ; vậy có thể coi nó bằng trung bình của các giá trị thứ 4 và thứ 5 và bằng  $(2,86+3,05)/2=2,96$ .

c) Do các giá trị mẫu khác nhau nên không có ước lượng mẫu của mốt.

## *Chương 5*

# **KIỂM NGHIỆM CÁC GIẢ THUYẾT THỐNG KÊ**

### **5.1. KHÁI NIỆM VỀ VIỆC KIỂM NGHIỆM CÁC GIẢ THUYẾT THỐNG KÊ**

#### **5.1.1. Ngẫu nhiên và bản chất**

Ta xét một bài toán thực tế: Có hai phương pháp sản xuất. Kết quả trong việc lấy mẫu cho biết là để sản xuất ra một sản phẩm theo phương pháp 1 phải cần:

2,0; 2,7; 2,5; 2,9; 2,3; 2,6 đơn vị nguyên liệu

và phương pháp 2 phải cần:

2,5 ; 3,2 ; 3,5 ; 3,8 ; 3,5 đơn vị nguyên liệu.

Như vậy một vấn đề đặt ra là sự khác nhau giữa 2 dãy số nói trên do bản chất là chính hay chỉ do ngẫu nhiên? Tất nhiên các câu hỏi tương tự như vậy cũng có thể đặt ra ở nhiều phần khác của lý thuyết thống kê.

#### **5.1.2. Giả thuyết thống kê và qui tắc kiểm nghiệm giả thuyết thống kê**

Để giải quyết vấn đề trên người ta nêu giả thuyết:

“Số trung bình các đơn vị nguyên liệu cần thiết để sản xuất ra một sản phẩm của hai phương pháp là bằng nhau”.

Và cần kiểm nghiệm xem giả thuyết ấy đúng hay sai.

Bài toán kiểm nghiệm giả thuyết thống kê tổng quát được đặt dưới dạng sau:



Cho đại lượng ngẫu nhiên  $X$  và một giả thuyết  $H_0$  về phân phối xác suất của  $X$ . Một mệnh đề khác với  $H_0$  được gọi là đối thuyết  $H_1$ . Cần kiểm nghiệm xem  $H_0$  đúng hay  $H_1$  đúng trên cơ sở mẫu lấy được là  $(X_1, X_2, \dots, X_n)$ . Muốn vậy ta lập không gian mẫu  $(X_1, X_2, \dots, X_n)$  và trên không gian mẫu ta xác định một miền  $W$  gọi là miền bác bỏ giả thuyết  $H_0$ , phần bù của  $W$  ký hiệu  $\bar{W}$  gọi là miền nhận giả thuyết  $H_0$ .

Mẫu lấy được  $(X_1, X_2, \dots, X_n)$  là một điểm xác định của không gian mẫu (vì các  $X_i, i=1, 2, \dots$  là các hằng số).

Nếu điểm  $(X_1, X_2, \dots, X_n) \in W$  thì ta coi giả thuyết  $H_0$  là sai và bác bỏ giả thuyết đó.

Nếu điểm  $(X_1, X_2, \dots, X_n) \in \bar{W}$  thì ta coi giả thuyết  $H_0$  là đúng và chấp nhận giả thuyết đó.

Như vậy một miền  $W$  xác định một qui tắc kiểm nghiệm giả thuyết và được gọi là qui tắc  $W$ . Trong việc chọn một qui tắc có thể mắc các sai lầm:

**Sai lầm loại 1:** Là sai lầm mắc phải khi giả thuyết  $H_0$  đúng mà ta bỏ giả thuyết ấy đi. Sai lầm này được đặc trưng bởi:

$$P(W / H_0) \quad (1)$$

Ký hiệu này chỉ xác suất bỏ  $H_0$  khi nó đúng; đó chính là xác suất của miền  $W$  khi  $H_0$  đúng.

**Sai lầm loại 2:** Là sai lầm mắc phải khi giả thuyết  $H_0$  sai mà ta lại chấp nhận giả thuyết ấy. Sai lầm này được đặc trưng bởi:

$$P(\bar{W} / H_1) \quad (2)$$

Ký hiệu này chỉ xác suất chấp nhận  $H_0$  khi nó sai; đó chính là xác suất của miền  $\bar{W}$  khi  $H_1$  đúng hay khi  $H_0$  sai.

Hai loại sai lầm dẫn đến hai tác hại khác nhau.

Một qui tắc sẽ càng tốt nếu cả hai loại sai lầm càng bé. Song điều ấy khó thực hiện. Vì vậy người ta coi một qui tắc là chấp nhận được nếu sai lầm loại 1 nhỏ hơn một hằng số  $\alpha$  cho trước:

$$P(W / H_0) \leq \alpha \quad (3)$$

$\alpha$  được gọi là *mức của qui tắc* và người ta cố gắng chọn qui tắc sao cho sai lầm loại 2 bé hợp lý nhất. Từ nay trở đi ta hiểu qui tắc là qui tắc chấp nhận được. Một qui tắc có sai lầm loại 2 càng bé càng tốt, nghĩa là:

$$P(\overline{W} / H_1) = 1 - P(W / H_1) \quad (4)$$

với  $P(W / H_1)$  càng lớn càng tốt.

Qui tắc W được coi là mạnh hơn qui tắc V nếu:

$$\begin{aligned} P(W / H_0) &\leq P(V / H_0) \\ P(W / H_1) &\geq P(V / H_1) \end{aligned} \quad (5)$$

Qui tắc W được coi là mạnh nhất nếu nó mạnh hơn mọi qui tắc V bất kỳ. Để minh họa việc xây dựng một qui tắc ta xét bài toán kiểm nghiệm giả thuyết về phương sai của đại lượng ngẫu nhiên có phân phối chuẩn  $N(\mu, \sigma^2)$ ,  $\mu$  và  $\sigma^2$  đều chưa biết. Căn cứ vào mẫu  $(X_1, X_2, \dots, X_n)$  hãy kiểm nghiệm giả thuyết  $H_0: \sigma = \sigma_0$  với đối thuyết  $H_1: \sigma = \sigma_1$  ( $\sigma_1 > \sigma_0$ ),  $\sigma_1$  và  $\sigma_0$  là các số đã biết, với mức  $\alpha$ . Ta dùng đại lượng thống kê phụ:

$$U = \frac{(n-1)S^2}{\sigma_0^2} \quad (6)$$

Nếu  $H_0$  đúng thì U là đại lượng có phân phối  $\chi^2$  với  $(n-1)$  bậc tự do. Ta lập qui tắc W với miền bác bỏ  $H_0$ :

$$W = \{(X_1, X_2, \dots, X_n): u > \chi^2\} \quad (7)$$

Trong đó  $\chi^2$  xác định từ:

$$\int_{\chi^2}^{+\infty} f_{n-1}(t) dt \leq \alpha \quad (8)$$

$f_{n-1}(t)$  là mật độ phân phối  $\chi^2$  với  $(n-1)$  bậc tự do.

Miền W bác bỏ giả thuyết như trên còn được gọi là *miền tới hạn*, và trị số  $\chi^2$  được gọi là *trị số tới hạn* của qui tắc kiểm nghiệm giả thuyết.



### 5.1.3. Ví dụ

Xét đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$ . Từ mẫu cỡ số lượng 22:

3,1; 3,3; 2,9; 3,0; 3,1; 3,2; 2,8; 2,7

3,1; 3,2; 2,9; 3,0; 2,9; 3,1; 2,8; 2,9

3,2; 3,3; 2,9; 3,1; 3,2; 3,0.

Hãy kiểm nghiệm giả thuyết  $H_0: \sigma = 0,025$  với đối thuyết  $H_1: \sigma = 0,060$  với mức  $\sigma = 0,05$ . Theo (6), (7), (8), miền  $W$  xác định bởi:

$$W = \left\{ (X_1, X_2, \dots, X_n) : u = \frac{\sum_{i=1}^{22} (X_i - \bar{X})^2}{0,025} > 32,7 \right\} \quad (9)$$

ở đây theo bảng VI (phần phụ lục) ta có  $\chi_{0,05}^2 = 32,7$  với phân phối  $\chi^2$  có 21 bậc tự do.

Thay các trị số quan sát được vào vế trái của (9) ta có:

$$u_0 = \frac{0,636}{0,025} = 25,5 < 32,7$$

nghĩa là  $u_0 \in \bar{W}$ ; vậy giả thuyết  $H_0$  được chấp nhận.

## 5.2. KIỂM NGHIỆM GIẢ THUYẾT ĐƠN VÀ QUI TẮC MẠNH NHẤT

### 5.2.1. Định lý Nâyman-Piécson

Cho đại lượng ngẫu nhiên  $X$  có mật độ phân phối xác suất  $f(x, \theta)$ , dạng của  $f$  đã biết còn  $\theta$  chưa biết. Xét giả thuyết:

$$H_0: \theta = \theta_0$$

với đối thuyết:

$$H_1: \theta = \theta_1 \quad (10)$$

( $\theta_0$  và  $\theta_1$  đều đã biết). Trường hợp này ta có bài toán kiểm nghiệm giả thuyết đơn (Giả thuyết  $H$  được coi là đơn nếu theo  $H$  thì  $\theta_0$  trùng với

một điểm  $\theta_1$  đã biết; Giả thuyết  $H$  được coi là phức tạp nếu theo  $H$  thì  $\theta \in \omega$  với  $\omega$  là một miền nào đó chứa nhiều hơn một điểm). Ở đây giả thuyết  $H_0$  cũng như đối thuyết  $H_1$  đều là đơn.

Để kiểm nghiệm giả thuyết đơn với đối thuyết cũng đơn J. Nêyman và E. Piécson đã chứng minh rằng: tồn tại qui tắc mạnh nhất và chỉ ra cách xây dựng qui tắc ấy.

### **Định lý Nêyman-Piécson.**

Khi kiểm nghiệm giả thuyết  $H_0: \theta = \theta_0$  với đối thuyết  $H_1: \theta = \theta_1$ , nếu:

$$L(X_1, \dots, X_n, \theta_j) = \prod_{i=1}^n f(X_i, \theta_j) \quad (j = 0, 1)$$

thì tồn tại một qui tắc mạnh nhất  $W$  với miền bác bỏ giả thuyết  $H_0$ :

$$W = \{(X_1, \dots, X_n) : L(X_1, \dots, X_n, \theta_1) \geq C L(X_1, \dots, X_n, \theta_0)\} \quad (11)$$

trong đó  $C$  là một hằng số xác định bởi:

$$P(W / H_0) \leq \alpha \quad (12)$$

với  $\alpha$  là mức của qui tắc đã cho.

### **5.2.2. Ví dụ**

Cho đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$ ,  $\sigma^2$  đã biết còn  $\theta$  chưa biết. Hãy tìm qui tắc mạnh nhất để kiểm nghiệm giả thuyết  $H_0: \theta = \theta_0$  với đối thuyết  $H_1: \theta = \theta_1$  ( $\theta \geq \theta_1$ ), với mức  $\alpha$  ( $\alpha, \theta_0, \theta_1$  đã biết). Ở đây:

$$L(X_1, \dots, X_n, \theta_j) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_j)^2} \quad (j = 0, 1) \quad (13)$$

Theo (11) thì miền  $W$  được xác định bởi:

$$W = \left\{ (X_1, \dots, X_n) : e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_1)^2} \geq C e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0)^2} \right\}$$



Từ đó rút ra:

$$(\theta_1 - \theta_0) \sum_{i=1}^n X_i - n(\theta_1^2 - \theta_0^2) \geq \sigma^2 \ln C$$

vậy:

$$W = \left\{ (X_1, \dots, X_n) : \bar{X} \geq \frac{\sigma^2 \ln C + n(\theta_1^2 - \theta_0^2)}{n(\theta_1 - \theta_0)} = A \right\} \quad (14)$$

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  và  $A$  xác định từ:

$$P(\bar{X} \geq A \mid H_0) \leq \alpha \quad (15)$$

Theo (11), (14) và (15) thì tính chất mạnh nhất của  $W$  không phụ thuộc trị số cụ thể của  $\theta_1$ , mà chỉ cần điều kiện  $\theta_1 > \theta_0$ , vì thế qui tắc (14) cũng là qui tắc mạnh nhất để kiểm nghiệm giả thuyết  $H_0$  với đối thuyết phức tạp  $H_1: \theta > \theta_0$ , trong trường hợp này người ta gọi  $W$  là qui tắc mạnh đều nhất (đều với mọi  $\theta > \theta_0$ ).

Nếu  $\theta_1 < \theta_0$ , tương tự như trên, ta có qui tắc mạnh đều nhất để kiểm nghiệm  $H_0$  với đối thuyết  $H_1: \theta < \theta_0$ . Qui tắc ấy có miền bác bỏ  $H_0$  là:

$$W'' = \left\{ (X_1, \dots, X_n) : \bar{X} \leq A' \right\} \quad (16)$$

với:

$$P(\bar{X} < A' \mid H_0) \leq \alpha$$

### 5.3. MỘT SỐ QUI TẮC KIỂM NGHIỆM GIẢ THUYẾT THỐNG KÊ QUAN TRỌNG THƯỜNG DÙNG

#### 5.3.1. Qui tắc t (qui tắc Stiuđơn)

Qui tắc t dựa trên đại lượng thống kê có phân phối Stiuđơn (mục 3, chương 4). Dưới đây ta xét 2 bài toán để minh họa qui tắc t và việc áp dụng qui tắc đó.

**Bài toán 1:** Xét đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$  với  $\mu, \sigma^2$  đều không biết; lấy mẫu  $(X_1, \dots, X_n)$ . Hãy kiểm nghiệm giả thuyết  $H_0: \mu = \mu_0$  với đối thuyết  $H_1: \mu \neq \mu_0$ .

Ở đây ta không thể áp dụng ngay định lý Nâyman-Piécson được vì  $H_0$  và  $H_1$  rất phức tạp (vì  $\sigma$  tùy ý).

Để giải quyết bài toán này người ta dùng qui tắc Stiudon  $t$  có miền bác bỏ giả thuyết  $H_0$ :

$$W = \{(X_1, \dots, X_n) : |t| \geq t_\alpha\} \quad (17)$$

$$\text{với: } t = \frac{\bar{X} - \mu_0}{S} \sqrt{n}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S = \left\{ \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \right\}^{1/2}$$

xác định bởi:  $\int_{-t_\alpha}^{t_\alpha} f_{n-1}(t) dt = 1 - \alpha$ , trong đó  $f_{n-1}(t)$  là mật độ phân phối

Stiudon với  $(n-1)$  bậc tự do.

**Bài toán 2:** Cho đại lượng ngẫu nhiên  $X$  có phân phối chuẩn  $N(\mu_1, \sigma^2)$  và đại lượng ngẫu nhiên  $Y$  cũng có phân phối chuẩn  $N(\mu_2, \sigma^2)$ ;  $\mu_1, \mu_2$  và  $\sigma^2$  đều chưa biết. Ta lấy mẫu với số lượng  $N_1$  ứng với  $X$  và số lượng  $N_2$  ứng với  $Y$ . Hãy kiểm nghiệm giả thuyết  $H_0: \mu_1 = \mu_2$  với đối thuyết  $H_1: \mu_1 \neq \mu_2$  với mức  $\alpha$ . Trường hợp này qui tắc  $t$  có miền bác bỏ giả thuyết:

$$W = (X_1, \dots, X_{N_1}), (Y_1, \dots, Y_{N_2}) : |t| \geq t_\alpha$$

$$= \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}} \right| > t_\alpha \quad (18)$$

$t_\alpha$  được xác định từ:  $\int_{-t_\alpha}^{t_\alpha} f_{N_1+N_2-2}(t) dt = 1 - \alpha$ . Với  $f_{N_1+N_2-2}(t)$  là mật

độ của phân phối Stiudon với  $(N_1+N_2-2)$  bậc tự do.



Trường hợp phải kiểm nghiệm  $H_0$  với đôi thuyết  $H_1^1: \mu_1 > \mu_2$  thì khi ấy miền  $W_1$  bác bỏ giả thuyết  $H_0$  có dạng:

$$W_1 = \{(X_1, \dots, X_{N_1}), (Y_1, \dots, Y_{N_2}) : t > t_\alpha^1\} \quad (19)$$

với:  $t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}}$ .

và  $t_\alpha^1$  được xác định từ:  $\int_{t_\alpha^1}^{+\infty} f_{N_1+N_2-2}(t)dt = \alpha$ .

Trường hợp phải kiểm nghiệm  $H_0$  với đôi thuyết  $H_1^2: \mu_1 < \mu_2$  thì miền  $W_2$  bác bỏ giả thuyết  $H_0$  có dạng:

$$W_2 = \{(X_1, \dots, X_{N_1}), (Y_1, \dots, Y_{N_2}) : t < t_\alpha^2\}$$

với:  $\int_{-\infty}^{t_\alpha^2} f_{N_1+N_2-2}(t)dt = \alpha$ .

### 5.3.2. Qui tắc $\chi^2$

Qui tắc  $\chi^2$  dựa trên đại lượng thống kê có phân phối  $\chi^2$ . Qui tắc  $\chi^2$  là một trong những qui tắc kiểm nghiệm giả thuyết được biết sớm nhất do E. Piécson nêu ra. Sau đây ta xét một số bài toán áp dụng qui tắc đó.

**Bài toán 1:** Kiểm tra phân phối xác suất của một đại lượng ngẫu nhiên. Cho đại lượng ngẫu nhiên  $X$  và giả thuyết  $H_0$ : "Hàm phân phối của  $X$  có dạng  $F(x)$  đã biết".

Để kiểm nghiệm giả thuyết trên người ta lấy mẫu  $(X_1, \dots, X_n)$  rồi làm như sau: chia miền giá trị của  $X$  ra làm  $k$  phần. Gọi  $n_i$  là số lần trị số  $X_j$  ( $j=1, \dots, n$ ) rơi vào khoảng thứ  $i$ , rõ ràng là  $\sum_{i=1}^k n_i = n$  (sau khi lấy mẫu thì  $n_i$  hoàn toàn xác định).

Nếu  $X$  có hàm phân phối xác suất là  $F(x)$  đã biết thì có thể tính được xác suất để  $X$  lấy trị số ở khoảng thứ  $i$  là  $p_i$ , với  $n$  thí nghiệm độc lập có thể coi số phần tử  $X_j$  ( $j=1, \dots, n$ ) rơi vào khoảng thứ  $i$  là  $np_i$ .

Trị số  $(n_i - np_i)^2$  được dùng làm cơ sở để xem xét phân phối xác suất của  $X$  có gần  $F(x)$  hay không. E. Piécson đưa ra đại lượng thống kê:

$$Q^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (20)$$

Rõ ràng có thể coi phân phối xác suất của  $X$  càng gần  $F(x)$  khi  $Q^2$  càng bé. Có thể chứng minh rằng khi  $n \rightarrow +\infty$  thì phân phối của  $Q^2$  tiến tới phân phối  $\chi^2$  với  $k-r-1$  bậc tự do, trong đó  $r$  chỉ số tham số chưa biết của  $F(x)$  (các tham số này phải được ước lượng theo kết quả của mẫu bằng phương pháp hợp lý cực đại để thay vào  $F(x)$  thì mới tính được các  $p_i$  trong công thức (4)). Miền bác bỏ giả thuyết  $H_0$  của qui tắc  $\chi^2$  được xác định bởi:

$$W = \{(X_1, \dots, X_n) : Q^2 > \chi_{\alpha}^2\} \quad (21)$$

với:  $\int_{\chi_{\alpha}^2}^{+\infty} f_{k-r-1}(t) dt = \alpha$ .

$\alpha$  là mức của qui tắc,  $f_{k-r-1}(t)$  là mật độ phân phối với  $k-r-1$  bậc tự do.

Trong thực tế, để đảm bảo sai số được hợp lý, người ta thường chia khoảng và lấy số lượng mẫu sao cho các  $n_i$  đều lớn, chẳng hạn  $n_i \geq 5$ .

**Bài toán 2:** Xét tính độc lập của các đại lượng ngẫu nhiên.

Trong trường hợp này các kết quả lấy mẫu xếp vào một bảng theo hai tính chất  $A, B$  (xem bảng) số  $n_{ij}$  ghi vào ô  $(i,j)$  ứng với hàng  $i$ , cột  $j$  chỉ số lần xuất hiện sự kiện  $A_i \cap B_j$ .



		A						Tổng số
		$A_1$	$A_2$	...	$A_i$	...	$A_r$	
B	$B_1$	$n_{11}$	$n_{21}$	...	$n_{i1}$	...	$n_{r1}$	$n_{.1}$
	...	...	...	...	...	...	...	...
	$B_j$	$n_{1j}$	$n_{2j}$	...	$n_{ij}$	...	$n_{rj}$	$n_{.j}$
	...	...	...	...	...	...	...	...
	$B_s$	$n_{1s}$	$n_{2s}$	...	$n_{is}$	...	$n_{rs}$	$n_{.s}$
Tổng số		$n_{.1}$	$n_{.2}$		$n_{.i}$		$n_{.r}$	$n$

Theo bảng trên ta thấy  $\sum_{j=1}^s n_{ij} = n_{.i}$  (số lần xuất hiện  $A_i$ ) và

$\sum_{i=1}^r n_{ij} = n_{.j}$  (số lần xuất hiện  $B_j$ ) và  $\sum_{j=1}^s n_{.j} = n$ .

Nếu  $A_i$  và  $B_j$  độc lập thì phải có:

$$P(A_i \cap B_j) = P(A_i).P(B_j) = \frac{n_{.i}}{n} \times \frac{n_{.j}}{n}$$

Khi ấy với số lượng mẫu là  $n$  thì số lần xuất hiện  $A_i \cap B_j$  là:

$$n \times \frac{n_{.i}}{n} \times \frac{n_{.j}}{n} = \frac{n_{.i} \times n_{.j}}{n}$$

Ta lập:

$$v^2 = \sum_{i=1}^r \sum_{j=1}^s \left( \frac{n_{ij} - \frac{n_{.i} \times n_{.j}}{n}}{\frac{n_{.i} \times n_{.j}}{n}} \right)^2 \quad (22)$$

Có thể chứng minh rằng khi  $n \rightarrow +\infty$  thì phân phối xác suất của  $v^2$  tiến tới phân phối  $\chi^2$  với  $(r-1).(s-1)$  bậc tự do.

Để kiểm nghiệm giả thuyết  $H_0$  : “A và B độc lập với nhau” (hay các  $A_i$  và  $B_j$  độc lập với nhau) với mức  $\alpha$ , ta dùng qui tắc  $\chi^2$  có miền  $W$  bác bỏ giả thuyết  $H_0$  xác định bởi:

$$v^2 > \chi_{\alpha}^2 \quad (23)$$

với  $\chi^2_\alpha$  xác định từ:

$$\int_{\chi^2_\alpha}^{+\infty} f_{(r-1)(s-1)}(t) dt = \alpha$$

$f_{(r-1)(s-1)}(t)$  là hàm mật độ phân phối  $\chi^2$  với  $(r-1).(s-1)$  bậc tự do.

### 5.3.3. Quy tắc F

Dựa trên đại lượng thống kê có phân phối F người ta ứng dụng quy tắc F để giải quyết các bài toán như sau:

**Bài toán 1:** Xét xem 2 phương sai của 2 đại lượng ngẫu nhiên có phân phối chuẩn có bằng nhau hay không. Giả sử 2 đại lượng ngẫu nhiên X và Y có phân phối chuẩn với phương sai  $\sigma_1^2$  và  $\sigma_2^2$ , ứng với mẫu có số lượng  $n_1$  và  $n_2$ . Hãy kiểm nghiệm giả thuyết  $H_0: \sigma_1 = \sigma_2$  với đối thuyết  $H_1: \sigma_1 > \sigma_2$  với mức  $\alpha$ .

Ta dùng đại lượng thống kê:  $F = \frac{S_1^2}{S_2^2}$  (với  $S_i^2$  ( $i = 1, 2$ ) là phương sai mẫu).

(Ở đây giả sử từ mẫu đã lấy được ta có  $S_1^2 > S_2^2$  để tránh trường hợp  $\frac{S_1^2}{S_2^2}$  quá bé làm cho việc xây dựng quy tắc trở thành phức tạp).

Nếu  $H_0$  đúng thì F có phân phối Phisơ-Snêđêco với  $(n_1-1)$  và  $(n_2-1)$  bậc tự do. Ở đây quy tắc F có miền bác bỏ giả thuyết  $H_0$  xác định bởi:  $F > f_\alpha$ , với  $f_\alpha$  được xác định từ:  $P\{F > f_\alpha\} = \alpha$  khi  $H_0$  đúng.

**Bài toán 2:** (Một bài toán phân tích phương sai). Cho K đại lượng ngẫu nhiên có phân phối chuẩn  $N(\mu_i, \sigma^2)$  với các  $\mu_i$  ( $i=1, \dots, K$ ) và  $\sigma^2$  đều chưa biết. Hãy kiểm nghiệm giả thuyết  $H_0: \mu_1 = \dots = \mu_K$  với đối thuyết  $H_1$ : các  $\mu_i$  không bằng nhau, với mức  $\alpha$ .

Gọi  $X_{ij}$  là giá trị mẫu thứ j của đại lượng  $X_i$ , ta đặt:



$$\bar{X} = \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} X_{ij}$$

với  $n = \sum_{i=1}^K n_i$  (tất cả có  $n$  số liệu, với  $X_i$  có  $n_i$  số liệu).

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

Ta có thể chứng minh được:

$$\sum_{i=1}^K \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^K n_i (\bar{X}_i - \bar{X})^2 + \sum_{i=1}^K \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

và viết tắt:

$$Q = Q_1 + Q_2 \quad (24)$$

Ta thấy  $Q_1$  được sinh ra do sự khác nhau giữa các số  $X_{ij}$  trong nội bộ mỗi nhóm (ứng với đại lượng ngẫu nhiên  $X_i$ ) còn  $Q_2$  được sinh ra do sự khác nhau giữa các nhóm.

Để kiểm nghiệm  $H_0$  với đối thuyết  $H_1$  ta dùng qui tắc  $F$  có miễn bác bỏ giả thuyết  $H_0$ :

$$W = \{ \{ X_{ij} \} : F > f_\alpha \} \quad (25)$$

với:

$$F = \frac{\frac{Q_1}{K-1}}{\frac{Q_2}{n-K}} = \frac{\frac{1}{K-1} \sum_{i=1}^K (\bar{X}_i - \bar{X})^2 n_i}{\frac{1}{n-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2} \quad (26)$$

Theo (25) và (26) thì ta bác bỏ giả thuyết khi tỉ lệ giữa độ phân tán giữa các nhóm ( $Q_1$ ) với độ phân tán trong nội bộ các nhóm ( $Q_2$ ) được coi là lớn và điều này xảy ra khi các  $\mu_i$  không bằng nhau.

Hằng số  $f_\alpha$  ở (25) được xác định từ:

$$\int_{f_0}^{+\infty} f_{(K-1),(n-1)}(t) dt = \alpha$$

$f_{(K-1),(n-1)}(t)$  là mật độ của phân phối  $F$  với  $(K-1), (n-1)$  bậc tự do.

#### 5.3.4. Qui tắc về dấu

Cho 2 đại lượng ngẫu nhiên  $X$  và  $Y$  với mẫu  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Giả sử  $X$  và  $Y$  độc lập với nhau và có các hàm mật độ liên tục  $f(x), g(y)$ . Hãy kiểm nghiệm giả thuyết  $H_0: "f(x) = g(x)"$  ( $X$  và  $Y$  có cùng hàm mật độ như nhau) với đối thuyết  $H_1: "f(x) \neq g(x)"$ , với mức  $\alpha$ .

Qui tắc về dấu dựa trên cơ sở phân tích dấu của  $Z = X - Y$  ( $Z_i = X_i - Y_i$ ,  $i=1, 2, \dots, n$ ). Nếu  $H_0$  đúng thì có thể chứng minh:

$$P(Z > 0) = P(Z < 0) = \frac{1}{2} \quad (27)$$

Gọi  $D_n$  là số các hiệu  $(X_i - Y_i)$  mang dấu dương thì  $\frac{D_n}{n}$  là tần suất của sự kiện  $(Z > 0)$  qua  $n$  phép thử. Nếu  $H_0$  đúng thì  $\frac{D_n}{n} \approx \frac{1}{2}$  với  $n$  lớn.

Có thể chứng minh:

$$\begin{aligned} \left\{ \left| \frac{D_n}{n} - \frac{1}{2} \right| \geq a \right\} &= \left\{ D_n \leq \frac{n}{2} - na \right\} \\ \left\{ D_n \geq \frac{n}{2} + na \right\} &= \left\{ \min(D_n, A_n) \leq \frac{n}{2} - na \right\} \end{aligned} \quad (28)$$

Với  $A_n$  là số các hiệu  $(X_i - Y_i)$  mang dấu âm ( $A_n + D_n = n$ ). Theo (27) và theo công thức Béc-nu-li ta có thể tính xác suất (khi  $H_0$  đúng) của các sự kiện ở công thức (28).

Qui tắc về dấu có miền bác bỏ giả thuyết  $H_0$ :

$$W = \{(X_1, Y_1), \dots, (X_n, Y_n) : \min(D_n, A_n) \leq m_\alpha\} \quad (29)$$

với  $m_\alpha$  được xác định từ:

$$P(W / H_0) \leq \alpha \quad (30)$$



Các trị số  $m_u$  được ghi trong bảng IX.

Ta biết khi  $n$  khá lớn thì phân phối xác suất của  $D_n$  (hoặc  $A_n$ ) gần với  $N\left(\frac{1}{2}, \frac{n}{4}\right)$  nên trong thực tế với  $n > 30$  có thể xem  $D_n$  (hoặc  $A_n$ ) có phân phối chuẩn (khi  $H_0$  đúng).

### 5.3.5. Quy tắc về nhóm

Cũng như quy tắc về dấu, quy tắc về nhóm là quy tắc không để ý đến tham số (phi tham số) được áp dụng rộng rãi.

Người ta dùng quy tắc này để kiểm nghiệm xem các phần tử của một dãy số có được sắp một cách ngẫu nhiên không? Nói cách khác, mỗi phần tử có xuất hiện cùng một xác suất không?

Giả sử có các giá trị của 2 đại lượng ngẫu nhiên  $X$  và  $Y$  được sắp theo thứ tự sau:

$$XXXX Y XXXXX YYYYYY XX \quad (31)$$

Ta nói trong hàng (31) có 5 nhóm, mỗi nhóm chỉ chứa toàn  $X$  (hay toàn  $Y$ ) đứng liên tiếp nhau, ở đây có 3 nhóm  $X$  và 2 nhóm  $Y$ .

Nếu tất cả có  $n$  giá trị của  $X$  và của  $Y$  thì số nhóm ít nhất có thể là 2 và nhiều nhất có thể là  $n$ . Người ta chứng minh được:

**Định lý:** *Nếu các giá trị  $X$  và  $Y$  được sắp một cách ngẫu nhiên thì số nhóm  $u$  được thành lập từ  $n$  giá trị trong đó có  $n_1$  giá trị của  $X$ ,  $n_2$  giá trị của  $Y$  ( $n_1 + n_2 = n$ ) cũng là một đại lượng ngẫu nhiên với:*

$$P(U = u) = 2 \frac{C_{n_1-1}^{u-1} \cdot C_{n_2-1}^{u-1}}{C_n^{n_1}} \quad \text{nếu } u \text{ chẵn.} \quad (32)$$

$$P(U = u) = \frac{C_{n_1-1}^{u-3} \cdot C_{n_2-1}^{u-1} + C_{n_1-1}^{u-1} \cdot C_{n_2-1}^{u-3}}{C_n^{n_1}} \quad \text{nếu } u \text{ lẻ} \quad (33)$$

Bây giờ xét dãy số chỉ kết quả thí nghiệm. Ta nêu giả thuyết  $H_0$ : “Dãy số ấy được sắp một cách ngẫu nhiên”. (Ở đây và về sau  $H_1$  có nội dung trái với  $H_0$ , nghĩa là  $H_1$  đúng khi  $H_0$  sai và ngược lại, ta không cần nhắc lại nữa).

Để kiểm nghiệm  $H_0$  ta chia các dãy số thành 2 loại X và Y. Một số được gọi là X nếu nó bé hơn *trung vị* của dãy, các số không phải X thì được gọi là Y. Sắp xếp lại ta có một dãy như (31). (Sắp dãy theo thứ tự tăng dần, nếu có  $(2m + 1)$  số thì số thứ  $(m + 1)$  được gọi là trung vị).

Qui tắc nhóm có miền bác bỏ giả thuyết  $H_0$  xác định bởi:

$$U \leq u_{\frac{1}{2}\alpha} \text{ hoặc } U \geq u_{1-\frac{1}{2}\alpha}$$

với  $\alpha$  là mức của qui tắc và:

$$U_\varepsilon = \max\{u': P(U \leq u') \leq \varepsilon\} \text{ với } \varepsilon < 0,5$$

$$U_\varepsilon = \min\{u': P(U \leq u') \geq \varepsilon\} \text{ với } \varepsilon > 0,5$$

với xác suất tính theo công thức (32), (33).

## BÀI TẬP CHƯƠNG 5

**Bài 1:** Từ tập hợp chính có phân phối chuẩn  $N(0; 25)$  ta lấy mẫu có số lượng 9 với  $\bar{X} = 1,95$ . Hãy kiểm nghiệm giả thuyết  $H_0: \theta=2$  với đối thuyết  $H_1: \theta=3$  với mức  $\alpha=0,05$ .

**Giải:** theo công thức (14) ta có:

$$W = \{(X_1, \dots, X_n) : \bar{X} > A\}$$

và A được xác định từ:

$$P = \{\bar{X} > A\} = P\left\{\frac{\bar{X} - 2}{5} \sqrt{9} > \frac{A - 2}{5} \cdot 3 = A_1\right\} = 0,05$$

Theo bảng II thì  $A_1=1,65$  và  $A = \frac{1,65}{3} \times 5 + 2 = 4,75$ . Theo mẫu đã có thì  $\bar{X} = 1,95 < 4,75$  vậy  $X \in W$  và theo giả thuyết  $H_0$  được chấp nhận.



**Bài 2:** Từ tập hợp chính gồm những trục được sản xuất ra từ một máy tiện tự động người ta lấy mẫu gồm 10 cái rồi đo đường kính của chúng, so với yêu cầu ta có các sai số sau:

Thứ tự trục	1	2	3	4	5	6	7	8	9	10
Sai số	+2	+1	-2	+3	+2	+4	-2	+5	+3	+4

Giả sử các sai số trên tuân theo luật phân phối chuẩn  $N(\mu, \sigma^2)$  với  $\mu$  và  $\sigma^2$  đều chưa biết. Hãy kiểm nghiệm giả thuyết  $H_0: \mu = 0$  với đối thuyết  $H_1: \mu \neq 0$  với mức  $\alpha = 1\%$ .

**Giải:** Ở đây  $\bar{X} = +2$ ;  $S = 2,43$ ,  $t = \frac{2-0}{2,43} \sqrt{10} = 2,60$ .

Theo bảng VII với phân phối Stiuđon có 9 bậc tự do và  $\alpha = 1\%$  thì  $t_\alpha = 3,25$ . Ở đây  $t = 2,6 < t_\alpha = 3,25$ , vậy giả thuyết  $H_0$  được chấp nhận (với sai lầm loại 1 là 1% và sai lầm loại 2 là hợp lý).

**Bài 3:** Với dãy số trong bài toán mở đầu ở **5.1**, giả sử hai đại lượng ngẫu nhiên ứng với hai dãy số đều có phân phối chuẩn với phương sai bằng nhau. Cần kiểm nghiệm giả thuyết  $H_0: \mu_1 = \mu_2$  (ở đây giả thiết  $X$  ứng với dãy số thứ nhất và  $EX = \mu_1$ ,  $Y$  ứng với dãy số thứ 2 và  $EY = \mu_2$ ) với đối thuyết  $H_1: \mu_1 \neq \mu_2$ , với mức  $\alpha = 0,05$ .

**Giải:** Theo mẫu đã biết thì:

$$\bar{X} = 2,5; (N_1 - 1)S_1^2 = 0,5; \bar{Y} = 3,3; (N_2 - 1)S_2^2 = 0,78.$$

Theo công thức  $|t|$ , ta có:

$$|t| = \left| \frac{2,5 - 3,3}{\sqrt{0,5 + 0,78 \left( \frac{1}{5} + \frac{1}{6} \right)}} \right| = 3,48$$

Theo bảng 7 có:  $t_\alpha = 2,26$ . Ta thấy  $|t| > t_\alpha$ , vậy  $H_0$  bị bác bỏ.

**Bài 4:** Bóng đèn điện tử của một nhà máy sản xuất ra có thể làm việc được 500 giờ. Bây giờ người ta áp dụng một phương pháp mới để

sản xuất bóng đèn, rồi lấy mẫu gồm 100 cái đem thí nghiệm thì được kết quả: Giờ làm việc trung bình  $\bar{X} = 560$  giờ,  $\frac{S^2}{\sqrt{100}} = 45$  giờ. Hãy kiểm nghiệm giả thuyết  $H_0$  : “Phương pháp mới cũng như phương pháp cũ” với mức  $\alpha = 0,05$ .

**Giải:** Ở đây ta gặp lại trường hợp bài toán 1 của qui tắc t (qui tắc Stiuđon) trong 5.3, nhưng với n lớn (n=100).

Vẫn dùng miền bỏ giả thuyết:  $|t| > C_\alpha$  nhưng  $C_\alpha = C_{0,05}$  xác định theo hàm Laplat:  $C_{0,05} = 1,96$ . Từ mẫu ta có:

$$t = \frac{\bar{X} - 500}{\frac{S}{\sqrt{n}}} = \frac{560 - 500}{45} = 1,33$$

ở đây  $|t| < C_\alpha$ , vậy giả thuyết  $H_0$  được chấp nhận.

**Bài 5:** Nghiên cứu các lần hỏng hóc của một quá trình sản xuất ứng với 8 kíp làm việc thì được bảng (1).

Hãy kiểm nghiệm giả thuyết  $H_0$  : “Số lần hỏng hóc đối với mỗi kíp có xác suất như nhau” (nghĩa là số lần hỏng hóc phân phối đều đối với các kíp) với mức  $\alpha = 0,05$ .

**Bảng 1**

Kíp	Số lần hỏng hóc $n_i$	$np_i$	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	16	18	-2	0,22
2	17	18	-1	0,06
3	19	18	1	0,06
4	16	18	-2	0,22
5	24	18	6	2,00
6	19	18	1	0,06
7	17	18	-1	0,06
8	16	18	-2	0,22
$\Sigma$	144	144	0	2,90



**Giải:** ở đây  $n=144$ , ta tìm các  $np_i$ ,  $\frac{(n_i - np_i)^2}{np_i}$  theo bảng (1) và

được  $Q^2=2,90$ . Vì phân phối không chứa tham số chưa biết nên  $r=0$  và khi  $n \rightarrow +\infty$  ta có phân phối  $\chi^2$  với (23) bậc tự do, theo (21) và bảng VII thì  $\chi_{0,05}^2 = 14,1$ . Ở đây  $Q^2=2,9 < 14,1$  vậy  $H_0$  được chấp nhận.

**Bài 6:** Căn cứ vào kết quả lấy mẫu ghi ở bảng (2) hãy kiểm nghiệm giả thuyết  $H_0$ : "Đại lượng ngẫu nhiên  $X$  có phân phối chuẩn" với mức  $\alpha = 0,05$ .

**Giải:** Để tìm các  $p_i$  cần ước lượng kỳ vọng  $\mu$  và phương sai  $\sigma^2$  của  $X$ . giả sử  $X$  có phân phối chuẩn  $N(\mu, \sigma^2)$ , theo phương pháp hợp lý cực đại ta có:  $\hat{\mu} = \bar{X}$  và  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  với  $\hat{\mu}$  và  $\hat{\sigma}^2$  là các ước lượng của  $\mu$  và  $\sigma^2$  mà ta sẽ dùng thay cho  $\mu$  và  $\sigma^2$ .

Theo bảng (2) có  $\hat{\mu} = 5,15$  và  $\hat{\sigma} = 0,76$ , vậy các  $p_i$  được tính từ  $N(5, 15; (0,76)^2)$ . Ví dụ ta tính  $p_2$ :

$$\begin{aligned} p_2 &= P\{3,6 < x < 4,2\} \\ &= P\left\{ \frac{3,6 - 5,15}{0,76} < \frac{x - \mu}{\sigma} < \frac{4,2 - 5,15}{0,76} \right\} \\ &= P\left\{ -2,04 < \frac{x - \mu}{\sigma} < -1,25 \right\} \\ &= \Phi(-1,25) - \Phi(-2,04) = 0,0849 \end{aligned}$$

**Bảng 2**

Trị số của $X$	Số kết quả
3,0-3,6	2
3,6-4,2	8
4,2-4,8	35
4,8-5,4	43
5,4-6,0	22
6,0-6,6	15
6,6-7,2	5
Cộng	130

( $\Phi$  là hàm Laplat).

Vậy:

$$np_2 = 130 \times 0,0849 = 11,04.$$

Tương tự:

$$np_2 = 2,09; np_3 = 28,24;$$

$$np_4 = 39,85; np_5 = 31,10;$$

$$np_6 = 13,43; np_7 = 3,65.$$

Vì vậy ta có:

$$Q^2 = \frac{(2 - 2,69)^2}{2,69} + \frac{(8 - 11,04)^2}{11,04} + \dots + \frac{(5 - 3,65)^2}{3,65} = 6,22$$

Vì có 2 tham số cần ước lượng nên ở đây phân phối  $\chi^2$  tương ứng có  $7 - 2 - 1 = 4$  bậc tự do. Ở đây miễn bác bỏ giả thuyết  $H_0$  là:

$$W = \{(X_1, \dots, X_n) : Q^2 > \chi_{0,05}^2 = 9,49\}$$

$6,22 < 9,49$ , vậy giả thuyết  $H_0$  được chấp nhận.

**Bài 7:** Jô-han-sen đo chiều dài của 12.000 quả đậu và phân ra 16 loại thì được bảng (3). Hãy kiểm nghiệm giả thuyết  $H_0$ : “Chiều dài các quả đậu tuân theo luật phân phối chuẩn” với mức  $\alpha = 0,1\%$ .

**Giải:** Cũng như ở bài 6, ta cần ước lượng kỳ vọng  $\mu$  và phương sai  $\sigma^2$  của  $X$  và được:

$$\hat{\mu} = 8,5012; \hat{\sigma}^2 = (0,6163)^2$$

(Thực ra để ước lượng  $\mu$  và  $\sigma^2$  ở đây có thể căn cứ ngay vào bảng (3) và có kể đến hiệu chỉnh Sippa).

Ở đây  $Q^2 = 196,5$  và  $\chi_{0,001}^2 = 34,5$  (ứng với  $\chi^2$  với 13 bậc tự do).  $196,5 > 34,5$ , vậy giả thuyết  $H_0$  bị bác bỏ, nghĩa là chiều dài của các quả đậu không tuân theo phân phối chuẩn.



**Bảng 3**

Thứ tự	Các lớp	Số lần xuất hiện $n_i$	$np_i$	$n_i - np_i$
1	bé hơn 7,00	32	68	-36
2	7,00-7,25	103	132	-29
3	7,25-7,50	239	310	-71
4	7,50-7,75	624	617	7
5	7,75-8,00	1187	1046	141
6	8,00-8,25	1650	1506	144
7	8,25-8,50	1883	1842	41
8	8,50-8,75	1930	1920	10
9	8,75-9,00	1638	1698	-60
10	9,00-9,25	1130	1277	-147
11	9,25-9,50	737	817	-80
12	9,50-9,75	427	444	-17
13	9,75-10,00	221	205	16
14	10,00-10,25	110	81	29
15	10,25-10,50	57	27	30
16	lớn hơn 10,50	32	10	22
<b>Tổng số</b>		<b>12.000</b>	<b>12.000</b>	<b>0</b>

**Bài 8:** Trong bảng (4) đưa ra phân phối màu tóc và lông mày của 46.542 người Thụy Điển được gọi nhập ngũ trong một khoảng thời gian nhất định. Hãy kiểm nghiệm giả thuyết  $H_0$  : “Màu tóc và màu lông mày độc lập với nhau” với mức  $\alpha = 0,05$ .

**Giải:** Áp dụng lý luận đã nêu trong qui tắc  $\chi^2$  5.3, với miễn bác bỏ giả thuyết  $H_0$ :

$$v^2 > \chi_{0,05}^2 = 3,8.$$

Vì ở đây  $(r-1).(s-1)=1$  nên  $\chi^2$  có một bậc tự do. Áp dụng công thức (22), ta được:

$$v^2 = 19,288$$

Vậy giả thuyết  $H_0$  bị bác bỏ.

**Bảng 4**

		Màu tóc		Tổng số
		Sáng hay hung	Tối	
Màu lông mày	Sáng hay hung	30.472	3.238	33.710
	Tối	3.364	9.468	12.832
Tổng cộng		33.836	12.706	46.542

**Bài 9:** Xét dãy số trong bài toán mở đầu ở 5.1, hãy kiểm nghiệm giả thuyết  $H_0: \sigma_1 = \sigma_2$  với đối thuyết  $H_1: \sigma_1 \neq \sigma_2$  với mức  $\alpha=0,05$ . ở đây ta giả sử các dãy số liệu từ 2 phương pháp đều theo luật phân phối chuẩn với phương sai là  $\sigma_1^2$  và  $\sigma_2^2$ .

**Giải:** Ta đặt  $F = \frac{S_2^2}{S_1^2}$  và miền bác bỏ giả thuyết  $H_0$  là:

$$W = \{(X_1, \dots, X_n) : F > f_{\alpha} = f_{0,05}\}$$

Tra bảng VIII từ phân phối F với (25) và (26) bậc tự do ta có  $f_{0,05} = 5,19$ . Theo mẫu ta có  $S_1^2 = 0,10$ ;  $S_2^2 = 0,195$  nên  $F = 1,95 < f_{\alpha}$  vậy giả thuyết  $H_0$  được chấp nhận.

**Bài 10:** Người ta sản xuất bóng điện từ 4 loại nguyên liệu khác nhau, kết quả thí nghiệm cho ở bảng (5), với giả thiết là giờ làm việc được của một bóng điện tuân theo luật phân phối chuẩn. Hãy kiểm nghiệm giả thuyết  $H_0$  : “ Các loại nguyên liệu không có ảnh hưởng đến chất lượng bóng điện” với mức  $\alpha=0,05$ .



**Bảng 5**

Nguyên liệu	Giờ làm việc của bóng điện							
I	1600	1610	1650	1680	1700	1700	1800	1820
II	1580	1640	1640	1700	1750			
III	1460	1550	1600	1620	1640	1660	1740	
IV	1510	1520	1530	1570	1600	1680		

**Giải:** Ta có:

$$\bar{X}_1 = 1677; \bar{X}_2 = 1662; \bar{X}_3 = 1636; \bar{X}_4 = 1568$$

$$\frac{Q_1}{K-1} = \frac{44,361}{3} = 14,787$$

$$\frac{Q_2}{n-K} = \frac{151,351}{22} = 6,880$$

và:

$$F = \frac{14,787}{6,880} = 2,15$$

Theo phân phối F với (25) bậc tự do có  $f_{\alpha} = 3,05$  ở đây  $F < f_{0,05}$ , vậy  $H_0$  được chấp nhận.

**Bài 11:** Chọn ngẫu nhiên 15 gia đình công nhân ở Vacsava mà mỗi gia đình có 5 người, rồi hỏi xem mỗi gia đình đã tiêu bao nhiêu tiền vào các sinh hoạt văn hoá và bao nhiêu tiền về việc đi lại (đi lại bằng tàu, xe...) thì được bảng (6). Hãy kiểm nghiệm giả thuyết  $H_0$ : “Chi phí về sinh hoạt văn hoá bằng chi phí đi lại” với đối thuyết  $H_1$ : “Hai chi phí có giá trị khác nhau”, với mức  $\alpha = 0,05$ .

**Giải:** Theo mẫu  $\min(D_n; \Lambda_n) = \min(13; 2) = 2$ , theo bảng IX  $m_{0,05} = 3$  vậy  $H_0$  bị bác bỏ.

**Bảng 6**

Thứ tự	Chi phí về sinh hoạt văn hoá	Chi phí về đi lại	Dấu của hiệu
1	12,1	5,3	+
2	6,2	5,7	+
3	6,0	4,2	+
4	8,4	4,6	+
5	4,2	10,8	-
6	6,8	5,0	+
7	6,9	5,2	+
8	20,0	12,7	+
9	13,5	6,3	+
10	7,2	3,2	+
11	8,5	5,0	+
12	10,7	4,8	+
13	7,9	5,7	+
14	1,3	8,4	-
15	6,9	6,7	+

**Bài 12:** Trong một khu vực nghỉ mát người ta thống kê số người du lịch đã đến đây trong 10 mùa hè liên thì được các con số (nghìn người): 8,3; 8,5; 8,5; 8,0; 8,6; 8,7; 9,0; 9,1; 9,2; 8,2.

Hãy kiểm nghiệm giả thuyết  $H_0$  : “ Các số được đặt một cách ngẫu nhiên” (nói cách khác số người đến nghỉ mát không có chiều hướng thay đổi) với mức  $\alpha = 0,05$  .

**Giải:** Trung vị của dãy trên là 8,55, ta có dãy:

XXXX YYYYYY X

Số nhóm  $u=3$ , theo bảng X với  $n_1=5$ ,  $n_2=5$  thì  $U_{0,025}=2$ ,  $U_{0,975}=9$ . Ở đây  $2 < 3 < 9$ , vậy  $H_0$  được chấp nhận.



**Bài 13:** Có hai đơn vị công nhân cùng sản xuất ra một loại sản phẩm, đơn vị một được bổ túc về kỹ thuật, đơn vị 2 không được bổ túc về kỹ thuật. Năng suất của đơn vị 1: 18,3; 17,0; 18,1; 18,4; 18,7; 17,9; 18,6 và của đơn vị 2: 18,2; 16,4; 17,5; 18,0; 17,7.

Hãy kiểm nghiệm giả thuyết  $H_0$  : “Việc bổ túc kỹ thuật không có tác dụng” (hay: năng suất của 2 đơn vị như nhau, hay: các số trên được sắp một cách ngẫu nhiên) với mức  $\alpha = 0,05$  .

**Giải:** Sắp các số liệu theo thứ tự không giảm: 16,4; 17,0; 17,7; 17,9; 18,0; 18,1; 18,2; 18,3; 18,4; 18,6; 18,7. Gọi năng suất của đơn vị một là X, của đơn vị 2 là Y ta có dãy:

Y X Y Y X Y X Y X X X X

Số nhóm  $u=8$ ; theo bảng X  $U_{0,025}=3$ ,  $U_{0,975}=10,3 < 8 < 10$ , vậy giả thuyết  $H_0$  được chấp nhận.

## *Phần phụ lục*

### **Phụ lục 1: BỔ TÚC TOÁN**

#### **1. Hai qui tắc kết hợp**

a) Qui tắc kết hợp cộng: Nếu đối tượng x có a cách chọn và đối tượng y có b cách chọn thì ta sẽ có (a+b) cách chọn để được x hoặc y.

b) Qui tắc kết hợp nhân: Nếu đối tượng x có a cách chọn và đối tượng y có b cách chọn thì ta sẽ có (a.b) cách chọn để được x và y.

**2. Chỉnh hợp:** Chỉnh hợp chập k của n phần tử ( $k \leq n$ ) là những nhóm gồm k phần tử có thể lập được từ n phần tử mà giữa chúng hoặc khác nhau về thành phần (phần tử) hoặc khác nhau về thứ tự giữa các thành phần (phần tử) tạo nên chúng và ta ký hiệu số chỉnh hợp là  $A_n^k$ :

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

**3. Hoán vị:** Hoán vị của n phần tử là số cách sắp xếp n phần tử ấy mà giữa chúng chỉ khác nhau về thứ tự giữa các phần tử và ta ký hiệu số hoán vị là  $P_n$ :

$$P_n = n(n-1)\dots 2.1 = n!$$

(Chú ý:  $P_n = A_n^n$ ,  $0! = 1! = 1$ ).

**4. Tổ hợp:** Tổ hợp chập k của n phần tử ( $k \leq n$ ) là những nhóm gồm k phần tử có thể lập được từ n phần tử mà giữa chúng chỉ khác nhau về thành phần (phần tử) tạo nên chúng và ta ký hiệu số tổ hợp là  $C_n^k$ :

$$C_n^k = \frac{n!}{(n-k)!k!}$$

#### **5. Nhị thức Niuton:**

$$(x+a)^n = \sum_{k=0}^n C_n^k x^{n-k} a^k$$



## Phụ lục 2: CÁC BẢNG SỐ

Bảng I

Giá trị số của hàm số  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3914	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	9653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3929	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551

**Bảng 1 tiếp theo**

[illegible]



Giá trị của hàm Laplat:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12556	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20194	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32881	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408

Bảng II tiếp theo

x	0	1	2	3	4	5	6	7	8	9
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	48788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	0,49865		3,1	48903	3,2	49931	3,3	49952	3,4	49966
3,5	49977		3,6	49984	3,7	49989	3,8	49993	3,9	49995
4,0	499968									
4,5	499997									
5,0	49999997									



**Bảng III**

Các giá trị của hàm:  $P_k(a) = \frac{a^k e^{-a}}{k!}$

<b>k \ a</b>	<b>0,1</b>	<b>0,2</b>	<b>0,3</b>	<b>0,4</b>	<b>0,5</b>	<b>0,6</b>
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075817	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003

k \ a	0,7	0,8
0	0,496585	0,449329
1	0,347610	0,359463
2	0,121663	0,143785
3	0,028388	0,038343
4	0,004968	0,007669
5	0,000695	0,001227
6	0,000081	0,000164
7	0,000008	0,000019
8		0,000002
9		
10		
11		
12		
13		
14		
15		



0,9	1,0	2,0	3,0
0,406570	0,367877	0,135335	0,049787
0,365913	0,367879	0,270671	0,149361
0,164661	0,183940	0,270671	0,224042
0,049398	0,061313	0,180447	0,224042
0,011115	0,015328	0,090224	0,168031
0,002001	0,003066	0,036089	0,100819
0,000300	0,000511	0,012030	0,050409
0,000039	0,000073	0,003437	0,021604
0,000004	0,000009	0,000859	0,008101
	0,000001	0,000191	0,002701
		0,000038	0,000810
		0,000007	0,000221
		0,000001	0,000055
			0,000013
			0,000003
			0,000001

k \ a	4,0	5,0	
0	0,018316	0,006738	
1	0,073263	0,033690	
2	0,146525	0,084224	
3	0,195367	0,140374	
4	0,195367	0,175467	
5	0,156293	0,175467	
6	0,104194	0,146223	
7	0,059540	0,104445	
8	0,029770	0,065278	
9	0,013231	0,036266	
10	0,005292	0,018133	
11	0,001925	0,008242	
12	0,000642	0,003434	
13	0,000197	0,001321	
14	0,000056	0,000472	
15	0,000015	0,000157	



6,0	7,0	8,0	9,0
0,002479	0,000912	0,000335	0,000123
0,014873	0,006383	0,002684	0,001111
0,044618	0,022341	0,010735	0,004998
0,089235	0,052129	0,028626	0,014994
0,133853	0,091226	0,057252	0,033737
0,160625	0,127717	0,091604	0,060727
0,160623	0,149003	0,122138	0,091090
0,137677	0,149003	0,139587	0,117116
0,103258	0,130377	0,139587	0,131756
0,068838	0,101405	0,124077	0,131756
0,041303	0,070983	0,099262	0,118580
0,022529	0,045171	0,017219	0,097020
0,011262	0,026350	0,048127	0,072765
0,005199	0,014188	0,029616	0,050376
0,002228	0,007094	0,012924	0,032384
0,000891	0,003311	0,009026	0,019431

k \ a						
	4,0	5,0	6,0	7,0	8,0	9,0
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786
18		0,000004	1,000039	0,000232	0,000944	0,002893
19		0,000001	0,000012	0,000085	0,000397	0,001370
20			0,000004	0,000030	0,000159	0,000617
21			0,000001	0,000010	0,000061	0,000264
22				0,000003	0,000022	0,000108
23				0,000001	0,000008	0,000042
24					0,000003	0,000016
25					0,000001	0,000006
26						0,000002
27						0,000001



**Bảng IV**

Giá trị của hàm  $\sum_{m=0}^k \frac{a^m}{m!}$

$k \backslash a$	0,1	0,2	0,3
0	0,904837	0,818731	0,740818
1	0,995321	0,982471	0,963063
2	0,999845	0,998852	0,996400
3	0,999996	0,999943	0,999734
4	1,000000	0,999998	0,999984
5	1,000000	1,000000	0,999999
6	1,000000	1,000000	1,000000
7	1,000000	1,000000	1,000000

$$\frac{m e^{-a}}{m !}$$

0,4	0,5	0,6
0,670320	0,606531	0,548812
0,938448	0,909796	0,878099
0,992074	0,985612	0,976885
0,999224	0,998248	0,996642
0,999939	0,999828	0,999606
0,999996	0,999986	0,999962
1,000000	0,999999	0,999997
1,000000	1,000000	1,000000



k \ a						
	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,844195	0,808792	0,772483	0,735759	0,406006	0,199148
2	0,965858	0,952577	0,937144	0,919699	0,676677	0,423190
3	0,994246	0,990920	0,986542	0,981012	0,857124	0,647232
4	0,999214	0,998589	0,997657	0,996340	0,947348	0,815263
5	0,999909	0,999816	0,999958	0,999406	0,983437	0,916082
6	0,999990	0,999980	0,999958	0,999917	0,995467	0,966491
7	0,999998	0,999999	0,999997	0,999990	0,998904	0,988095
8	1,000000	1,000000	1,000000	0,999999	0,999763	0,996196
9				1,000000	0,999954	0,998897
10					0,999992	0,999707
11					0,999999	0,999928
12					1,000000	0,999983
13						0,999996
14						0,999999
15						1,000000

k \ a	4,0	5,0
0	0,018316	0,006738
1	0,091579	0,040428
2	0,238105	0,124652
3	0,433473	0,265026
4	0,628839	0,440493
5	0,785132	0,615960
6	0,889326	0,762183
7	0,948866	0,866628
8	0,978636	0,931806
9	0,991867	0,968172
10	0,997159	0,986305
11	0,999084	0,984547
12	0,999726	0,997981
13	0,999923	0,999202



6,0	7,0	8,0	9,0
0,002479	0,000912	0,000335	0,000123
0,017352	0,007295	0,003019	0,001234
0,061970	0,029636	0,013754	0,006232
0,151205	0,081765	0,042380	0,021228
0,285058	0,172991	0,099632	0,054963
0,445681	0,300708	0,191236	0,115690
0,606304	0,449711	0,313374	0,206780
0,743981	0,598714	0,452961	0,323896
0,847239	0,729091	0,592548	0,455652
0,916077	0,830496	0,716625	0,587408
0,957380	0,901479	0,815887	0,705988
0,979909	0,946650	0,888077	0,803008
0,991173	0,973000	0,936204	0,875773
0,996372	0,987188	0,965820	0,926149

k \ a	4,0	5,0
14	0,999979	0,999774
15	0,999994	0,999931
16	0,999998	0,999980
17	0,999999	0,999994
18	0,999999	0,999998
19	0,999999	0,999998
20	1,000000	0,999999
21		1,000000
22		
23		
24		
25		
26		
27		
28		



6,0	7,0	8,0	9,0
0,998600	0,994282	0,982744	0,058533
0,999491	0,997593	0,991770	0,977964
0,999825	0,999041	0,996283	0,988894
0,999943	0,999637	0,998407	0,994680
0,999982	0,999869	0,999351	0,997573
0,999994	0,999955	0,999748	0,998943
0,999998	0,999985	0,999907	0,999560
0,999999	0,999995	0,999967	0,999824
0,999999	0,999998	0,999989	0,999932
1,000000	0,999999	0,999997	0,999974
	0,999999	0,999999	0,999990
	1,000000	0,999999	0,999996
		1,000000	0,999998
			0,999999
			1,000000

**Bảng V**

**Một phần bảng các số ngẫu nhiên của KADURÔP**

3393	6270	4228	6069	9407	1865	8549	3217	2351	8410
9108	2330	2157	7416	0398	6173	1703	8132	9065	6717
7891	3590	2502	5945	3402	0491	4328	2365	6175	7695
9085	6307	6910	9174	1753	1797	9229	3422	9861	8357
2638	2908	6368	0398	5495	3283	0031	5955	6544	3883
1313	8330	0623	8600	4950	5414	7131	0134	7241	0651
3897	4202	3814	3505	1599	1649	2784	1994	5775	1406
4380	9543	1646	2850	8415	9120	8062	2421	6161	4634
1618	6309	7909	0874	0401	4301	4517	9197	3380	0434
4858	4676	7363	9141	6133	0549	1092	3461	7116	1496
5354	9142	0847	5393	5416	6505	7156	5634	9703	6221
0905	6986	9396	3975	9255	0537	2479	4589	0562	5345
1420	0470	8679	2328	3939	1292	0406	5428	3789	2882
3218	9080	6604	1813	8209	7039	2086	3369	4437	3798
9697	8431	4387	0622	6893	8788	2320	9358	5904	9539
0912	4964	0502	9683	4636	2861	2876	1273	7870	2030
4636	7072	4868	0604	3894	7182	8417	2367	7032	0003
2515	4734	9878	6761	5636	2949	3979	8650	3430	0635
5964	0412	5012	2369	6461	0678	3693	2928	3740	8047
7848	1523	7904	1521	1455	7089	8094	9872	0898	7174
5192	2571	3643	0707	3434	6818	5729	8614	4298	4129
8438	8325	9886	1805	0226	2510	3675	5058	2515	2388
8166	6349	0319	5436	6838	2460	6433	0644	7428	8556
9158	8263	6504	2562	1160	1256	1816	0690	1215	9590
6061	3525	4048	0382	4224	7148	8259	6526	5340	4064



2215	6716	0176	7252	7362
0594	6677	4277	5312	9797
2537	0650	5050	0838	1868
8186	6897	9319	8452	0781
9810	4151	1302	4064	5397
8101	3288	8794	6632	1275
7102	3728	6105	3900	7778
4919	7877	7875	9568	1967
3087	5020	9626	0297	1020
0308	6808	5926	5287	0757
1666	4644	7873	5759	1931
5200	8994	7981	5227	1910
5997	5551	8698	6322	2079
5094	5432	1067	8922	5909
2424	2394	1666	1562	5058
9096	7615	0264	1638	0543
2180	7723	7882	3109	5740
5496	3741	3667	5732	3000
9815	2164	6735	8960	1080
8458	6454	3571	3332	3690
9541	0812	0753	9456	6945
9220	2718	2215	0650	6249
3271	2196	1505	6897	2636
1704	0664	9425	4151	6825
8659	9807	3789	8110	8213
2182	1666	7373	4982	2368
1549	8441	3351	3079	0026
5118	4796	7035	2010	3449
6848	3420	6583	7520	4809



**Bảng V tiếp theo**

7988	0340	4740	9958	3951
0195	4773	8368	4190	1226
4562	7910	4780	3305	5591
3957	5828	4597	3837	5648
0163	8005	1704	7584	8098
4104	1253	4268	9849	4428
7402	2163	4597	3977	4311
6554	0086	2863	3259	9357
5425	9894	4126	5566	8815
1570	2555	1607	2272	3574
9136	0537	5600	3957	7263
1458	4331	9555	4483	4152
0158	0154	1124	8923	8253
6021	7623	2107	5760	8631
9662	1702	6044	4689	6547
3464	2397	2029	3949	9953
6742	3607	1382	4273	2461
6324	5383	1839	7732	5854
2555	1290	2427	3379	1289
1040	5670	4057	1880	8429
0349	7808	2446	5744	2542
1074	7385	7285	0876	9936
0969	0739	7120	2497	6977
1131	4928	7378	6629	8323
5904	2732	4430	2464	5291
2613	9025	2836	8493	8207
4161	6224	4184	2633	2736
7061	3847	9508	4528	1226
1575	3209	9070	0644	3614



3309	4853	4021	8644	3980
2050	3603	1812	4020	6573
6817	6736	4591	9037	2949
8310	3044	0433	1322	7664
7050	7670	1848	5173	2146
5637	5325	9367	5939	3191
8251	3537	5139	5050	1516
6129	7391	0429	2836	5284
1311	4847	2317	3561	4533
5151	2477	3250	7859	1764
4245	4289	9301	4788	3937
3395	7157	1084	0561	8388
6542	9757	7896	7029	4986
8068	1543	2019	3678	9248
3370	3763	4713	4726	7010
9366	3691	3221	3010	5948
2166	2583	6878	3080	6318
6844	3518	1638	7438	6106
1479	7233	3827	9134	1023
6342	6329	1364	1015	3892
0485	7584	6030	2040	8568
0315	9870	0611	0799	3905
6970	4028	2192	6139	4041
7669	5252	7613	7180	5878
3030	4044	2628	4619	3946
0820	8374	3852	3137	6565
6713	9803	9650	5060	7665

*Bảng V tiếp theo*

5318	1959	4783	1810	2020
0312	7739	9374	4788	4350
7406	4238	4279	6206	1699
3310	2487	3926	2233	8260
1289	8504	0911	2001	5804
5930	3361	6743	5995	4194
8792	5513	5583	6103	9872
5485	2643	4035	3089	4991
6655	7354	1903	6909	9776
1590	3309	2555	9334	3869
9122	9301	4741	0683	0253
7575	7513	0431	9659	8204
6632	8145	7080	3287	3336
2452	1224	0260	4158	3735
3736	2412	9066	8663	5408
5659	4145	2452	3340	5540
6494	8057	7173	5753	3592
2268	6361	7008	7995	8010
6356	5108	3033	3546	0827
5611	7390	2421	9873	9796
3539	6208	9450	7613	4591
3138	3326	7949	3667	5243
4796	2730	1020	2413	9291
8760	3502	5051	9024	3413
4144	0733	9224	5937	1801
7427	1125	5148	2787	8948
3907	7169	8117	4134	4040



4159	9180	1702	3121	2651
9205	2198	9098	1824	7501
0642	1412	6797	9138	5612
5811	9590	3383	0477	5168
4043	8583	9226	8380	1588
9044	5859	2998	3765	7890
4119	6651	6116	5813	7584
3071	5011	4724	3490	3027
2117	7644	9041	1179	4972
4011	3379	9603	7710	8469
3283	9340	7677	1030	6631
0983	5962	5702	0111	3408
9896	0924	3658	1160	2912
9080	2770	9375	3482	7891
2331	3275	2357	7034	6931
1273	9284	0351	2933	0369
2525	3418	9975	1270	8196
3612	5917	0848	9251	2687
2128	8715	7380	9909	5242
3617	2695	6030	4897	1861
5597	7749	1911	5101	1952
9108	4777	3450	1479	6380
8677	1526	9362	4810	8831
4857	4364	5740	4365	8179
1218	2074	5893	2599	3795
4606	2484	5415	2139	4324

**Bảng V tiếp theo**

1743	4469	6599	3945	4423
7387	2085	9028	0523	3488
5225	4842	7099	1096	1691
4079	7351	6174	6420	6520
6424	1467	2570	1911	2554
9768	2748	7783	6226	1929
3624	1827	6561	4344	3502
9436	1063	8385	7194	0991
2985	4648	4880	1792	3136
8415	0195	5304	9145	2952
7848	5611	4655	3600	3247
8431	2390	2577	8429	2124
9604	4299	1437	9609	7960
4243	7769	1416	3290	4908
4191	3326	2359	8606	5618
5071	3027	8864	5538	0402
1222	2070	5888	6520	7531
3028	3311	4212	3678	8329
5295	2501	6973	0151	5918
9772	6807	2032	8889	9478
3707	8153	7010	3760	0202
7954	9798	4191	3643	7982
5217	8019	6031	5441	9141
9430	5192	6962	5076	1500
9082	0540	3343	5061	2512
6881	6976	1020	1056	2442



2713	6196	4336	4778	9366
7818	7151	5126	0262	4002
2940	7567	9389	1771	6315
1204	4917	3690	3581	6603
3457	7112	6741	6919	2574
7639	8567	8636	2255	0671
4526	3143	3336	7305	8358
6871	7606	9020	1501	1826
6527	9712	1078	5129	3277
3512	3182	2816	1030	1096
9162	4092	6975	2917	7182
0957	1243	0173	4674	4658
5453	7065	3998	2753	1332
4628	1351	5809	9384	9755
0723	8419	9258	8300	1735
5858	9158	5360	9242	3447
4340	6877	3777	7248	7661
2535	7180	5967	8560	3721

*Bảng V tiếp theo*

1376	2045	3051	7128	9056
1722	0022	6020	4107	2635
1685	6284	6440	5387	1199
4221	2338	1295	9815	6337
8136	7378	6057	5580	3284
0039	0960	2904	5720	5479
8905	9605	7747	2493	0203
2010	9814	5467	9106	3272
2022	4562	6324	2071	1995
6232	9570	2758	0640	0424
2646	1115	8900	1428	1092
1202	5891	1582	4123	6758
8627	6732	7472	8158	8083
1443	4080	1033	6467	5358
2040	4708	4354	6072	6092
0996	4362	3136	9927	4320
9281	1532	5010	8392	2044
1516	2448	1226	8101	3195



**Bảng VI**

n \ $\alpha$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50
1.	0,000	0,001	0,004	0,016	0,064	0,148	0,455
2.	0,020	0,040	0,103	0,211	0,446	0,713	1,386
3.	0,115	0,185	0,352	0,584	1,005	1,424	2,37
4.	0,297	0,429	0,711	1,064	1,649	2,20	3,36
5.	0,554	0,752	1,145	1,610	2,34	3,00	4,35
6.	0,872	1,134	1,635	2,20	3,07	3,83	5,35
7.	1,239	1,564	2,17	2,83	3,82	4,67	6,35
8.	1,646	2,03	2,73	3,49	4,59	5,53	7,34
9.	2,09	2,53	3,32	4,17	5,38	6,39	8,34
10.	2,56	3,06	3,94	4,86	6,18	7,27	9,34
11.	3,05	3,61	4,58	5,58	6,99	8,15	10,34
12.	3,57	4,18	5,23	6,30	7,81	9,03	11,34
13.	4,11	4,76	5,89	7,04	8,63	9,93	12,34
14.	4,66	5,37	6,57	7,79	9,47	10,82	13,34
15.	5,23	5,98	7,26	8,55	10,31	11,72	14,34
16.	5,81	6,61	7,96	9,31	11,15	12,62	15,34

n \ $\alpha$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50
17.	6,41	7,26	8,67	10,08	12,00	13,53	16,34
18.	7,02	7,91	9,39	10,86	12,86	14,44	17,34
19.	7,63	8,57	10,11	11,65	13,72	15,35	18,34
20.	8,26	9,24	10,85	12,44	14,58	16,27	19,34
21.	8,90	9,92	11,59	13,24	15,44	17,18	20,3
22.	9,54	10,60	12,34	14,04	16,31	18,10	21,3
23.	10,20	11,29	13,09	14,85	17,19	19,02	22,3
24.	10,86	11,99	13,85	15,66	18,06	19,94	23,3
25.	11,52	12,70	14,61	16,47	18,94	20,9	24,3
26.	12,20	13,41	15,38	17,29	19,82	21,8	25,3
27.	12,88	14,12	16,15	18,11	20,7	22,7	26,3
28.	13,56	14,85	16,93	18,94	21,6	23,6	27,3
29.	14,26	15,57	17,71	19,77	22,5	24,6	28,3
30.	14,95	16,31	18,49	20,6	23,4	25,5	29,3



n	$\alpha$	0,30	0,20
1.		1,074	1,642
2.		2,41	3,22
3.		3,66	4,64
4.		4,88	5,99
5.		6,06	7,29
6.		7,23	8,56
7.		8,38	9,80
8.		9,52	11,03
9.		10,66	12,24
10.		11,78	13,44
11.		12,90	14,63
12.		14,01	15,81
13.		15,12	16,98
14.		16,22	18,15
15.		17,32	19,31
16.		18,42	20,5

	0,10	0,05	0,02	0,01	0,001
	2,71	3,84	5,41	6,64	10,83
	4,60	5,99	7,82	9,21	13,82
	6,25	7,82	9,84	11,34	16,27
	7,78	9,49	11,67	13,28	18,46
	9,24	11,07	13,39	15,09	20,5
	10,64	12,59	15,03	16,81	22,5
	12,02	14,07	16,62	18,48	24,3
	13,36	15,51	18,17	20,1	26,1
	14,68	16,92	19,68	21,7	27,9
	15,99	18,31	21,2	23,2	29,6
	17,28	19,68	22,6	24,7	31,5
	18,55	21,0	24,1	26,2	32,9
	19,81	22,4	25,5	27,7	34,6
	21,1	23,7	26,9	29,1	36,3
	22,3	25,0	28,3	30,6	37,7
	23,5	26,3	29,6	32,0	39,3



<div> <div><math>\alpha</math></div> <div>n</div> </div>	0,30	
17.	19,51	
18.	20,6	
19.	21,7	
20.	22,8	
21.	23,9	
22.	24,9	
23.	26,0	
24.	27,1	
25.	28,2	
26.	29,2	
27.	30,3	
28.	31,4	
29.	32,5	
30.	33,5	

0,20	0,10	0,05	0,02	0,01	0,001
21,6	24,8	27,6	31,0	33,4	40,8
22,8	26,0	29,8	32,3	34,8	42,3
23,9	27,2	30,1	33,7	36,2	43,8
25,0	28,4	31,4	35,0	37,6	45,3
26,2	29,6	32,7	36,3	38,9	46,8
27,3	30,8	33,9	37,7	40,3	48,3
28,4	32,0	35,2	39,0	41,6	49,7
29,6	33,2	36,4	40,3	43,0	51,2
30,7	34,4	37,7	41,7	44,3	52,6
31,8	35,6	38,9	42,9	45,6	54,1
32,9	36,7	40,1	44,1	47,0	55,5
34,0	37,9	41,3	45,4	48,3	56,9
35,1	39,1	42,6	46,4	49,6	58,3
36,2	40,3	43,8	48,0	50,0	59,7



**Bảng VII**  
**Các trị số  $t_{\alpha}$  được xác định từ đại lượng ngẫu nhiên  $T_k$**   
**có phân bố student với k bậc tự do**

$$P(|T_K| > t_{\alpha}) = \alpha \quad 2 \frac{\Gamma(k+1)}{\Gamma\left(\frac{k}{2}\right) \sqrt{\pi k}} \int_{t_{\alpha}}^{\infty} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}} dt = \alpha$$

$\alpha \backslash k$	0,05	0,02	0,01	0,005	0,001
1	12,706	31,821	63,657		636,6
2	4,303	6,965	9,925	14,089	31,600
3	3,182	4,541	5,841	7,453	12,922
4	2,776	3,747	4,604	5,597	8,610
5	2,571	3,365	4,032	4,773	6,869
6	2,447	3,143	3,707	4,317	5,959
7	2,365	2,998	3,499	4,029	5,408
8	2,306	2,896	3,355	3,833	5,041
9	2,262	2,821	3,250	3,690	4,781
10	2,228	2,764	3,169	3,581	4,587
12	2,179	2,681	3,055	3,428	4,318
14	2,145	2,624	2,977	3,326	4,140
16	2,120	2,583	2,921	3,252	4,015
18	2,101	2,552	2,878	3,193	3,922
20	2,086	2,528	2,845	3,153	3,849
22	2,074	2,508	2,819	3,119	3,792
24	2,064	2,492	2,797	3,092	3,745
26	2,056	2,479	2,779	3,067	3,704
28	2,048	2,467	2,763	3,047	3,674
30	2,042	2,457	2,750	3,030	3,646
$+\infty$	1,960	2,326	2,576	2,807	3,291

**Bảng VIII**

Các trị số  $f_{\alpha}$  được xác định từ đại lượng ngẫu nhiên  $F = \frac{S_1^2}{S_2^2}$  có phân phối F

với  $k_1$  và  $k_2$  bậc tự do theo công thức:  $P(E > f_{\alpha}) = \alpha$

A)  $\alpha = 0,05$

		$k_1$ (bậc tự do của tử $S_1^2$ )															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$k_2$ (bậc tự do của mẫu $S_2^2$ )	1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246	246
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70



	1	2	3	4	5	6	7
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83
14	4,69	3,74	3,34	3,11	2,96	2,85	2,76
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49
22	4,30	3,44	3,05	2,82	2,66	2,65	2,46
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42

**Bảng VIII tiếp theo**

8	9	10	11	12	13	14	15	16
2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60
2,77	2,71	2,67	2,66	2,60	2,58	2,55	2,53	2,51
2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44
2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38
2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33
2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29
2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25
2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21
2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18
2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,16
2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,13
2,38	2,32	2,27	2,23	2,20	2,18	2,15	2,13	2,11
2,36	2,30	2,25	2,21	2,18	2,15	2,13	2,11	2,09



**Bảng VIII (Tiếp theo từ cột thứ 16 ở phần trên)**

	17	18	19	20	22	24	26	28	30	40	50	60	80	100
1	247	247	248	248	249	249	249	250	250	251	252	252	252	253
2	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,68	8,67	8,67	8,66	8,65	8,64	8,63	8,62	8,62	8,59	8,58	8,57	8,56	8,55
4	5,83	5,82	5,81	5,80	5,79	5,77	5,76	5,75	5,75	5,72	5,70	5,69	5,67	5,66
5	4,59	4,58	4,57	4,56	4,54	4,53	4,52	4,50	4,50	4,46	4,44	4,43	4,41	4,41
6	3,91	3,90	3,88	3,87	3,86	3,84	3,83	3,82	3,81	3,77	3,75	3,74	3,72	3,71
7	3,48	3,47	3,46	3,44	3,43	3,41	3,40	3,39	3,38	3,34	3,32	3,30	3,29	3,27
8	3,19	3,10	3,16	3,15	3,13	3,12	3,10	3,09	3,08	3,04	3,02	3,01	2,99	2,97
9	2,97	2,96	2,95	2,94	2,92	2,90	2,89	2,87	2,86	2,83	2,80	2,79	2,77	2,76
10	2,81	2,80	2,78	2,77	2,75	2,74	2,72	2,71	2,70	2,66	2,64	2,62	2,60	2,59
11	2,69	2,67	2,66	2,65	2,63	2,61	2,59	2,58	2,57	2,53	2,51	2,49	2,47	2,46
12	2,58	2,57	2,56	2,54	2,52	2,51	2,49	2,48	2,47	2,43	2,40	2,38	2,36	2,35
13	2,50	2,48	2,47	2,46	2,44	2,42	2,41	2,39	2,38	2,34	2,31	2,30	2,27	2,26
14	2,43	2,41	2,40	2,39	2,37	2,35	2,33	2,32	2,31	2,27	2,24	2,22	2,20	2,19
15	2,37	2,35	2,34	2,33	2,31	2,29	2,27	2,26	2,25	2,20	2,18	2,16	2,14	2,12
16	2,32	2,30	2,29	2,28	2,25	2,24	2,22	2,21	2,19	2,15	2,12	2,11	2,08	2,07
17	2,27	2,26	2,24	2,23	2,21	2,19	2,17	2,16	2,15	2,10	2,08	2,06	2,03	2,02
18	2,23	2,22	2,20	2,19	2,17	2,15	2,13	2,12	2,11	2,06	2,04	2,02	1,99	1,98
19	2,20	2,18	2,17	2,16	2,13	2,11	2,10	2,08	2,07	2,03	2,00	1,98	1,96	1,94
20	2,17	2,15	2,14	2,12	2,10	2,08	2,07	2,05	2,04	1,99	1,97	1,95	1,92	1,91
21	2,14	2,12	2,11	2,10	2,07	2,05	2,04	2,02	2,01	1,96	1,94	1,92	1,88	1,86
22	2,11	2,10	2,08	2,07	2,05	2,03	2,01	2,00	1,98	1,94	1,91	1,89	1,86	1,85
23	2,09	2,07	2,06	2,05	2,02	2,00	1,99	1,97	1,96	1,91	1,88	1,86	1,84	1,82
24	2,07	2,05	2,04	2,03	2,00	1,98	1,97	1,95	1,94	1,89	1,86	1,84	1,82	1,80

Tiếp bảng VIII (Tiếp theo phần A với  $\alpha=0,05$ )

		$k_1$ (bậc tự do của tử $S_0^2$ )															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$k_2$ (bậc tự do của mẫu $S_1^2$ )	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,07
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07	2,05
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06	2,04
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	2,02
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03	2,01
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,99
	32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99	1,97
	34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97	1,95
	36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95	1,93
	38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94	1,92
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90
	42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99	1,96	1,93	1,91	1,89
	44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90	1,88
	46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97	1,94	1,91	1,89	1,87
	48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88	1,86
	50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87	1,85
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,82
	70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,79
	80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,77
	90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78	1,76
	100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,75
	125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83	1,80	1,77	1,75	1,72
	150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82	1,79	1,76	1,73	1,71
	200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72	1,69
	300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78	1,75	1,72	1,70	1,68
	500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,69	1,66
	$\geq 1000$	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76	1,73	1,70	1,68	1,65



**Bảng VIII** (Tiếp theo từ cột thứ 16 ở phần trên)

	17	18	19	20	22	24	26	28	30	40	50	60	80	100
25	2,05	2,04	2,02	2,01	1,98	1,96	1,95	1,93	1,92	1,87	1,84	1,82	1,80	1,78
26	2,03	2,02	2,00	1,99	1,97	1,95	1,93	1,91	1,90	1,85	1,82	1,80	1,78	1,76
27	2,02	2,00	1,99	1,97	1,95	1,93	1,91	1,90	1,88	1,84	1,81	1,79	1,76	1,74
28	2,00	1,99	1,97	1,96	1,95	1,91	1,90	1,88	1,87	1,82	1,79	1,77	1,74	1,73
29	1,99	1,97	1,96	1,94	1,92	1,90	1,88	1,87	1,85	1,81	1,77	1,75	1,73	1,71
30	1,98	1,96	1,95	1,93	1,91	1,89	1,87	1,85	1,84	1,79	1,76	1,74	1,71	1,70
32	1,95	1,94	1,92	1,91	1,88	1,86	1,85	1,83	1,82	1,77	1,74	1,71	1,69	1,67
34	1,93	1,92	1,90	1,89	1,86	1,84	1,82	1,80	1,80	1,75	1,71	1,69	1,66	1,65
36	1,92	1,90	1,88	1,87	1,85	1,82	1,81	1,79	1,78	1,73	1,69	1,67	1,64	1,62
38	1,90	1,88	1,87	1,85	1,83	1,81	1,79	1,77	1,76	1,71	1,68	1,65	1,62	1,61
40	1,89	1,87	1,85	1,84	1,81	1,79	1,77	1,76	1,74	1,69	1,66	1,64	1,61	1,59
42	1,87	1,86	1,84	1,83	1,80	1,78	1,76	1,74	1,73	1,68	1,65	1,62	1,59	1,57
44	1,86	1,84	1,83	1,81	1,79	1,77	1,75	1,73	1,72	1,67	1,63	1,61	1,58	1,56
46	1,85	1,83	1,82	1,80	1,78	1,76	1,74	1,72	1,71	1,65	1,62	1,60	1,57	1,55
48	1,84	1,82	1,81	1,79	1,77	1,75	1,73	1,71	1,70	1,64	1,61	1,59	1,56	1,54
50	1,83	1,81	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,63	1,60	1,58	1,54	1,52
60	1,80	1,78	1,76	1,75	1,72	1,70	1,68	1,66	1,65	1,59	1,56	1,53	1,50	1,48
70	1,77	1,75	1,74	1,72	1,70	1,67	1,65	1,64	1,62	1,57	1,53	1,50	1,47	1,45
80	1,75	1,73	1,72	1,70	1,68	1,65	1,63	1,62	1,60	1,54	1,51	1,48	1,45	1,43
90	1,74	1,72	1,70	1,69	1,66	1,64	1,62	1,60	1,59	1,53	1,49	1,46	1,43	1,41
100	1,73	1,71	1,69	1,68	1,65	1,63	1,61	1,59	1,57	1,52	1,48	1,45	1,41	1,39
125	1,70	1,69	1,67	1,65	1,63	1,60	1,58	1,57	1,55	1,49	1,45	1,42	1,39	1,36
150	1,69	1,67	1,66	1,64	1,61	1,59	1,57	1,55	1,53	1,48	1,44	1,41	1,37	1,34
200	1,67	1,66	1,64	1,62	1,60	1,57	1,55	1,53	1,52	1,46	1,41	1,39	1,35	1,32
300	1,66	1,64	1,62	1,61	1,58	1,55	1,53	1,51	1,50	1,43	1,39	1,36	1,32	1,30
500	1,64	1,62	1,61	1,59	1,56	1,54	1,52	1,50	1,48	1,42	1,38	1,34	1,30	1,28
≥1000	1,63	1,61	1,60	1,58	1,55	1,53	1,51	1,49	1,47	1,41	1,36	1,33	1,29	1,26

Bảng VIII

B)  $\alpha=0,01$ 

		$k_1$ (bậc tự do của tử $S_1^2$ )															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$k_2$ (bậc tự do của mẫu $S_2^2$ )	2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
	3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0	26,9	26,9	26,8
	4	21,2	18,0	16,1	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,4	14,3	14,2	14,2	14,2
	5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82	9,77	9,72	9,68
	6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52
	7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,27
	8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48
	9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,00	4,96	4,92
	10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97
	13	9,07	6,70	5,74	5,21	4,86	4,65	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78
	14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62
	15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37
	17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19
	19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12
	20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05
	21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03	2,99
	22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98	2,94
	23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93	2,89
	24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89	2,85
	25	7,77	5,57	4,48	4,18	3,86	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85	2,81
	26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90	2,86	2,82	2,78



**Bảng VIII (Tiếp theo từ cột thứ 16 ở phần trên)**

	17	18	19	20	22	24	26	28	30	40	50	60	80	100
$k_2$ (bậc tự do của mẫu $s_2^2$ )														
2	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5
3	26,8	26,8	26,7	26,7	26,6	26,6	26,6	26,5	26,5	26,4	26,4	26,3	26,3	26,2
4	14,1	14,1	14,0	14,0	14,0	13,9	13,9	13,9	13,8	13,7	13,7	13,7	13,6	13,6
5	9,94	9,61	9,58	9,55	9,51	9,47	9,43	9,40	9,38	9,29	9,24	9,20	9,16	9,13
6	7,48	7,45	7,42	7,40	7,35	7,31	7,28	7,25	7,23	7,14	7,09	7,06	7,01	6,99
7	6,24	6,21	6,18	6,16	6,11	6,07	6,04	6,02	5,99	5,91	5,86	5,82	5,78	5,75
8	5,44	5,41	5,38	5,36	5,32	5,28	5,25	5,22	5,20	5,12	5,07	5,03	4,99	4,96
9	4,89	4,86	4,83	4,81	4,77	4,73	4,70	4,67	4,65	4,57	4,52	4,48	4,44	4,42
10	4,49	4,46	4,43	4,41	4,36	4,33	4,30	4,27	4,25	4,17	4,12	4,08	4,04	4,01
11	4,18	4,15	4,12	4,10	4,06	4,02	3,99	3,96	3,94	3,86	3,81	3,78	3,73	3,71
12	3,94	3,91	3,88	3,86	3,82	3,78	3,75	3,72	3,70	3,62	3,57	3,54	3,49	2,47
13	3,75	3,72	3,69	3,66	3,62	3,59	3,56	3,53	3,51	3,43	3,38	3,34	3,30	3,27
14	3,59	3,56	3,53	3,51	3,46	3,43	3,40	3,57	3,35	3,27	3,22	3,18	3,14	3,11
15	3,45	3,42	3,40	3,37	3,33	3,29	3,26	3,24	3,21	3,13	3,08	3,05	3,00	3,98
16	3,34	3,31	3,58	3,26	3,22	3,18	3,15	3,12	3,10	3,02	2,97	2,93	2,89	2,86
17	3,24	3,21	3,18	3,16	3,12	3,08	3,05	3,03	3,00	2,92	2,87	2,83	2,79	2,76
18	3,16	3,13	3,10	3,08	3,03	3,00	3,00	2,96	2,92	2,84	2,78	2,75	2,70	2,68
19	3,08	3,05	3,03	3,00	2,96	2,92	2,89	2,87	2,84	2,76	2,71	2,67	2,63	2,60
20	3,02	2,99	2,96	2,94	2,90	2,86	2,83	2,80	2,78	2,69	2,64	2,61	2,56	2,54
21	2,96	2,93	2,90	2,88	2,84	2,80	2,77	2,74	2,72	2,64	2,58	2,55	2,50	2,48
22	2,91	2,88	2,85	2,83	2,78	2,75	2,72	2,69	2,67	2,58	2,53	2,50	2,45	2,42
23	2,86	2,83	2,80	2,78	2,74	2,70	2,67	2,64	2,62	2,54	2,48	2,45	2,40	2,37
24	2,82	2,79	2,76	2,74	2,70	2,66	2,63	2,60	2,58	2,49	2,44	2,40	2,36	2,33
25	2,78	2,75	2,72	2,70	2,66	2,62	2,59	2,56	2,54	2,45	2,40	2,36	2,32	2,29
26	2,71	2,72	2,69	2,66	2,62	2,58	2,55	2,53	2,50	2,42	2,36	2,33	2,28	2,25



Tiếp bảng VIII (Tiếp theo phần B với  $\alpha=0,01$ )

$k_1$ (bậc tự do của tử $S_1^2$ )																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87	2,82	2,78	2,75
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84	2,79	2,75	2,72
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81	2,77	2,73	2,69
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70	2,66
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80	2,74	2,70	2,66	2,62
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76	2,70	2,66	2,62	2,58
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72	2,67	2,62	2,58	2,54
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69	2,64	2,59	2,55	2,51
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52	2,48
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64	2,59	2,54	2,50	2,46
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62	2,56	2,52	2,47	2,44
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60	2,54	2,50	2,45	2,42
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,72	2,64	2,58	2,53	2,48	2,44	2,40
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70	2,63	2,56	2,51	2,46	2,42	2,38
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53	2,47	2,42	2,38	2,34
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35	2,31
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31	2,27
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27	2,23
90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33	2,29	2,24	2,21
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,26	2,22	2,19
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33	2,28	2,23	2,19	2,15
150	6,81	4,75	3,92	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31	2,25	2,20	2,16	2,12
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22	2,17	2,13	2,09
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24	2,19	2,14	2,10	2,06
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22	2,17	2,12	2,07	2,04
$\geq 1000$	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20	2,15	2,10	2,06	2,02



	17	18	19	20
27	2,71	2,68	2,66	2,63
28	2,68	2,65	2,63	2,60
29	2,66	2,63	2,60	2,57
30	2,63	2,60	2,57	2,55
32	2,58	2,55	2,53	2,50
34	2,55	2,51	2,49	2,46
36	2,51	2,48	2,45	2,43
38	2,48	2,45	2,42	2,40
40	2,45	2,42	2,39	2,37
42	2,43	2,40	2,37	2,34
44	2,40	2,37	2,35	2,32
46	2,38	2,36	2,33	2,30
48	2,37	2,33	2,31	2,28
50	2,35	2,32	2,29	2,27
55	2,31	2,28	2,25	2,23
60	2,28	2,25	2,22	2,20
70	2,23	2,20	2,18	2,15
80	2,20	2,17	2,14	2,12
90	2,17	2,14	2,11	2,09
100	2,15	2,12	2,09	2,07
125	2,11	2,08	2,05	2,03
150	2,09	2,06	2,03	2,00
200	2,06	2,02	2,00	1,97
300	2,03	1,99	1,97	1,94
500	2,00	1,97	1,94	1,92
≥1000	1,98	1,55,	1,92	1,90



**Bảng VIII** (Tiếp theo từ cột thứ 16 ở phần trên)

22	24	26	28	30	32	34	36	80	100
2,59	2,55	2,52	2,49	2,47	2,38	2,33	2,29	2,25	2,22
2,56	2,52	2,49	2,46	2,44	2,35	2,30	2,26	2,22	2,19
2,53	2,49	2,46	2,44	2,41	2,33	2,27	2,23	2,19	2,16
2,51	2,47	2,44	2,41	2,39	2,30	2,25	2,21	2,16	2,13
2,46	2,42	2,39	2,36	2,34	2,25	2,20	2,16	2,11	2,08
2,42	2,38	2,35	2,32	2,30	2,21	2,16	2,12	2,07	2,04
2,38	2,35	2,32	2,29	2,26	2,17	2,12	2,08	2,03	2,00
2,35	2,32	2,28	2,26	2,23	2,14	2,09	2,05	2,00	1,97
2,33	2,29	2,26	2,23	2,20	2,11	2,06	2,02	1,97	1,94
2,30	2,26	2,23	2,20	2,18	2,09	2,03	1,99	1,94	1,91
2,28	2,24	2,21	2,18	2,15	2,06	2,01	1,97	1,92	1,89
2,26	2,22	2,19	2,16	2,13	2,04	1,99	1,95	1,90	1,86
2,24	2,20	2,17	2,14	2,12	2,02	1,97	1,93	1,88	1,84
2,22	2,18	2,15	2,12	2,10	2,01	1,95	1,91	1,86	1,82
2,18	2,15	2,11	2,08	2,06	1,97	1,91	1,87	1,81	1,78
2,15	2,12	2,08	2,05	2,03	1,94	1,88	1,84	1,78	1,75
2,11	2,07	2,03	2,01	1,98	1,89	1,83	1,78	1,73	1,70
2,07	2,03	2,00	1,97	1,94	1,85	1,79	1,75	1,69	1,66
2,04	2,00	1,97	1,94	1,92	1,82	1,76	1,72	1,66	1,62
2,02	1,98	1,94	1,92	1,89	1,80	1,73	1,69	1,63	1,60
1,98	1,94	1,91	1,88	1,85	1,76	1,69	1,65	1,59	1,55
1,96	1,92	1,88	1,85	1,83	1,73	1,66	1,62	1,56	1,52
1,93	1,89	1,85	1,82	1,79	1,69	1,63	1,58	1,52	1,48
1,89	1,85	1,82	1,79	1,76	1,66	1,59	1,55	1,48	1,44
1,87	1,83	1,79	1,76	1,74	1,63	1,56	1,52	1,45	1,40
1,85	1,81	1,77	1,74	1,72	1,61	1,54	1,50	1,43	1,38



**Bảng IX**  
 Các số nguyên  $m_\alpha$  được xác định từ:  $P\{\min(D_n, A_n) \leq m_\alpha\} \leq \alpha$

n	$\alpha$		n	$\alpha$		n	$\alpha$	
	0,01	0,05		0,01	0,05		0,01	0,05
1			37	10	12	73	25	27
2			38	10	12	74	25	28
3			39	11	12	75	25	28
4			40	11	13	76	26	28
5			41	11	13	77	26	29
6		0	42	12	14	78	27	29
7		0	43	12	14	79	27	30
8	0	0	44	12	15	80	28	30
9	0	1	45	13	15	81	28	31
10	0	1	46	13	15	82	28	31
11	0	1	47	14	16	83	29	32
12	1	2	48	14	16	84	29	32
13	1	2	49	15	17	85	30	32
14	1	2	50	15	17	86	30	33
15	2	3	51	15	18	87	31	33
16	2	3	52	16	18	88	31	34
17	2	4	53	16	18	89	31	34
18	3	4	54	17	19	90	32	35

n	$\alpha$		n
	0,01	0,05	
19	3	4	55
20	3	5	56
21	4	5	57
22	4	5	58
23	4	6	59
24	5	6	60
25	5	7	61
26	6	7	62
27	6	7	63
28	6	8	64
29	7	8	65
30	7	9	66
31	7	9	67
32	8	9	68
33	8	10	69
34	9	10	70
35	9	11	71
36	9	11	72



$\alpha$		n	$\alpha$	
0,01	0,05		0,01	0,05
17	19			
18	20			
18	20			
18	21			
19	21			
19	21			
20	22			
20	22			
21	23			
21	23			
21	24			
22	24			
22	25			
22	25			
23	25			
23	26			
24	26			
24	27			

**Bảng X**  
**Các trị số tới hạn của quy tắc về nhóm.**

 $u_{0,025}$ 

$n_1$	$n_2$																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4																			
5			2	2															
6		2	2	3	3														
7		2	2	3	3	3													
8		2	3	3	3	4	4												
9		2	3	3	4	4	5	5											
10		2	3	3	4	5	5	5	6										
11		2	3	4	4	5	5	6	6	7									
12	2	2	3	4	4	5	6	6	7	7	7								
13	2	2	3	4	5	5	6	6	7	7	8	8							
14	2	2	3	4	5	5	6	7	7	8	8	9	9						
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10					
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11				
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11			
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12		
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14



# Bảng X tiếp theo

$U_{0.975}$

$n_1$	$n_2$																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2	4	6	8																	
3	5	7	8	9																
4	5	7	8	9																
5	5	7	8	9																
6	5	7	8	9																
7	5	7	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
8	5	7	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
9	5	7	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
10	5	7	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
11	5	7	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
12	5	7	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
13	5	7	9	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
14	5	7	9	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
15	5	7	9	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
16	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	25	26	27		
17	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	25	26	27		
18	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	25	26	27		
19	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	25	26	27		
20	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	25	26	27		

## TÀI LIỆU THAM KHẢO

1. “ *Cơ sở lý thuyết xác suất và thống kê toán*”  
(Tập thể tác giả) - Nhà xuất bản ĐH và THCN, Hà Nội-1972.
2. Trần Tuấn Điệp, Lý Hoàng Tú.  
“ *Giáo trình lý thuyết xác suất và thống kê toán*”.  
Nhà xuất bản ĐH và THCN, Hà Nội-1979.
3. Tống Đình Quý. “ *Hướng dẫn giải bài tập xác suất thống kê*”. Nhà  
xuất bản Giáo Dục Hà Nội- 1998.
4. Barnes J.W.  
“*Statistical analysis for engineers and Scientists*”.  
Mc. Graw-Hill. 1994.
5. Rodanôv Iu.A  
“*Bài giảng lý thuyết xác suất*”.  
M. 1968 (Tiếng Nga).
6. Khudxôn D.  
“*Thống kê cho các nhà vật lý*”.  
M. 1970 (Tiếng Nga).



## MỤC LỤC

<b>LỜI NÓI ĐẦU</b>	<b>3</b>
<b>Chương 1</b>	
<b>NHỮNG KHÁI NIỆM CƠ BẢN CỦA LÝ THUYẾT XÁC SUẤT</b>	<b>5</b>
1.1. Khái niệm về phép thử, biến cố và xác suất của biến cố	5
1.2. Các định nghĩa về xác suất của biến cố	7
1.3. Mối quan hệ và các phép tính giữa các biến cố	12
1.4. Xác suất của tổng hai biến cố xung khắc.	14
1.5. Xác suất có điều kiện - Định lý nhân xác suất.	16
1.6. Xác suất của tổng hai biến cố không xung khắc	19
1.7. Độ tin cậy của một hệ thống thiết bị	21
<b>Bài tập chương 1.</b>	<b>25</b>
<b>Chương 2.</b>	
<b>CÔNG THỨC XÁC SUẤT TOÀN PHẦN VÀ CÔNG THỨC BÂYÉT.</b>	
<b>DÃY PHÉP THỬ ĐỘC LẬP.</b>	<b>36</b>
2.1. Công thức xác suất toàn phần.	36
2.2. Công thức Bâyét.	39
2.3. Công thức Béc-nu-li cho dãy phép thử độc lập.	41
2.4. Các định lý giới hạn Moa-Vơ-Rơ Laplat	43
2.5. Định lý giới hạn Poátxông	47
2.6. Số lần xuất hiện chắc nhất.	48
2.7. Định lý tổng quát về các phép thử lặp	50
<b>Bài tập chương 2.</b>	<b>52</b>

### **Chương 3.**

ĐẠI LƯỢNG NGẪU NHIÊN. CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN	60
---	----

3.1. Khái niệm về đại lượng ngẫu nhiên. Bảng phân phối xác suất. Hàm phân phối xác suất.	60
3.2. Kỳ vọng toán của đại lượng ngẫu nhiên	68
3.3. Phương sai của đại lượng ngẫu nhiên	73
3.4. Kỳ vọng toán và phương sai của một số luật phân phối thường gặp	77
3.5. Một và trung vị.	84
3.6. Mômen.	86
<i>Bài tập chương 3.</i>	88

### **Chương 4.**

MẪU THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ.	99
4.1. Không gian mẫu.	99
4.2. Phân phối mẫu và phân phối chính xác.	102
4.3. Phân phối xác suất đại lượng thống kê trên không gian mẫu.	104
4.4. Phân phối tiệm cận chuẩn của đại lượng thống kê.	111
4.5. Hàm ước lượng của một tham số chưa biết.	112
4.6. Ước lượng tham số theo phương pháp hợp lý cực đại.	120
4.7. Ước lượng tham số bằng khoảng tin cậy.	124
4.8. Ước lượng tham số với mẫu có số lượng lớn	127
<i>Bài tập chương 4.</i>	128

### **Chương 5:**

KIỂM NGHIỆM CÁC GIẢ THUYẾT THỐNG KÊ	136
5.1. Khái niệm về việc kiểm nghiệm các giả thuyết thống kê	136
5.2. Kiểm nghiệm giả thuyết đơn và qui tắc mạnh nhất	139
5.3. Một số qui tắc kiểm nghiệm giả thuyết thống kê	141
<i>Bài tập chương 5.</i>	150
PHỤ LỤC.	160
Tài liệu tham khảo.	196



**NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI**

16 Hàng Chuối - Hai Bà Trưng - Hà Nội

Điện thoại: (04) 9715011; (04) 9721544. Fax: (04) 9714899

Email: [nxb@vnu.edu.vn](mailto:nxb@vnu.edu.vn)

★ ★ ★

***Chịu trách nhiệm xuất bản:***

*Giám đốc:* PHÙNG QUỐC BẢO

*Tổng biên tập:* PHẠM THÀNH HƯNG

***Chịu trách nhiệm nội dung:***

Hội đồng nghiệm thu giáo trình

Trường ĐHKHTN – Đại học Quốc gia Hà Nội

*Người nhận xét:* GS.TSKH. NGUYỄN XUÂN HÂN

PGS.TSKH. NGUYỄN VĂN HÙNG

PGS.TS. LÊ VĂN TRỰC

*Biên tập:* VŨ BÁ MAI

*Biên tập tái bản:* NGUYỄN THỊ THUY

*Trình bày bìa:* NGỌC ANH

---

**LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC**

Mã số: 1K - 02050 - 03204

In 1000 cuốn, khổ 14,5 x 20,5 tại Nhà in Đại học Quốc gia Hà Nội

Số xuất bản: 9/113/XB - QLXB, ngày 10/2/2004. Số trích ngang: 64 KH/XB

In xong và nộp lưu chiểu quý II năm 2004.